


THE USE OF SUMMATED RATINGS  
IN FACULTY EVALUATION,

by

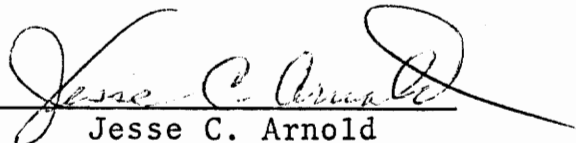
Edith Houston Carter,

Dissertation submitted to the Graduate Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of  
DOCTOR OF PHILOSOPHY  
in  
Educational Research

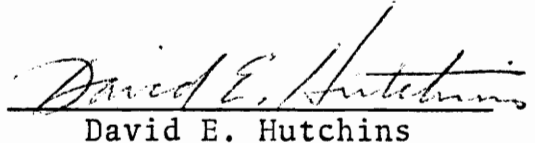
APPROVED:



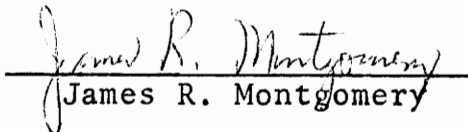
Robert B. Frary, Chairman



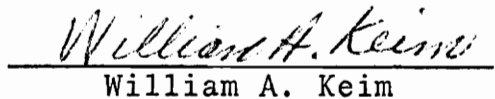
Jesse C. Arnold



David E. Hutchins



James R. Montgomery



William A. Keim

January, 1976

Blacksburg, Virginia

LD  
5655  
V856  
1976  
C37  
c.2

## ACKNOWLEDGMENTS

The writer wishes to express sincere appreciation to all of the people who gave very generously of their time and knowledge in support of the research for this study.

Special acknowledgments are extended to the following people: Dr. Robert Frary, my major professor, who provided incentive and encouragement from his very extensive knowledge in the field of Educational Research.

Dr. Jesse Arnold, chairman of the Department of Statistics, for giving me a growing knowledge and appreciation for the field of statistics.

Dr. James Montgomery, Director of Institutional Research, who provided a practical approach to developing an office of Institutional Research and who was always available to provide guidance and assistance.

Dr. David Hutchins, for encouragement to pursue the doctorate, and for the many opportunities to explore various areas of research.

Dr. William Keim, President, Pioneer Community College in Kansas City, Missouri, for incentive and encouragement in the field of Institutional Research and for his vast knowledge of institutional data needs.

Dr. Robert Sullins, President of New River Community College, for constant encouragement, inspiration and incentive needed to pursue the study.

Dr. Donald P. Hoyt, Director of Educational Research at Kansas State University, for his knowledge in the field of faculty evaluation and for very generously providing the necessary evaluation forms and manuals to conduct the study.

Dr. Fletcher Carter, my husband, for encouragement and assistance at home that provided the time needed to pursue the study.

Special appreciation is extended to the individual faculty members at New River Community College, Radford College, and Virginia Polytechnic Institute and State University for offering their support, assistance, and class time to conduct the study (participants in the study are listed on the following page).

Faculty Members Participating  
in the Study

<u>NRCC</u>	<u>RADFORD COLLEGE</u>	<u>VPI &amp; SU</u>
Edwin L. Barnes	David Albig	Richard Bambach
Irene P. Barr	Mary P. Balkus	Wilma Clay
James N. Bennett	Jack. R. Barnard	Richard C. Cole
A. William Claussen	Bennie E. Beeson, Jr.	Roderick Coleman
Gary C. Dudderar	Don Dwyer	Susan Geahart
Thomas Fagan	Barbara N. Ewell	Joseph A. Fry
Sennett W. Gore	Richard L. Hoffman	Judy Gardner
Helen M. Harvey	Authur A. Jannery	George R. Gross
Katherine H. Hillman	Myrl G. Jones	Ernest C. Houck
Betty F. Hines	Charlene Lutes	Bettibel Kreye
Bonita W. Leathers	George McConnell	Joseph Lamoureux
DeAnn B. Lineberry	John McCrimmon	Douglas McFarland
Melvin D. Mabry	Patrick B. Mikesell	Harlan B. Miller
Owen C. McKinnie, III	Raymond K. Moore	Robert W. Morrill
Alfred Nevola	Clifford C. Norse	Bruce E. Reed
Edith M. Ruben	Linda L. Poole	Clayton Peterson
Dennis M. Sayers	W. Keith Roberts	Barbara Price
Sharon R. Semones	Reginald J. Scolaro	Richard D. Shingles
Margaret L. Smith	G. Galen Stickley	Larry Springer
Mary C. Wesselkamper	Michael V. Vengrin	Karla Vollmar

## TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS . . . . .	ii
LIST OF TABLES . . . . .	vii
LIST OF FIGURES . . . . .	viii
 Chapter	
1. INTRODUCTION . . . . .	1
Statement of the Problem . . . . .	5
Need for the Study . . . . .	6
2. REVIEW OF THE LITERATURE . . . . .	8
Early History . . . . .	8
Uses of Student Evaluators of Instruction . . . . .	10
Criticisms of Methods & Use of Student Evaluation . . . . .	12
Student Input into Instructor Evaluation . .	14
Summated Ratings . . . . .	17
Standardized Rating Scales . . . . .	18
Summary . . . . .	21
3. RESEARCH DESIGN AND PROCEDURE . . . . .	22
Population and Sample . . . . .	24
Profile of Faculty Participating in the Study . . . . .	26
Procedures . . . . .	27
Processing of Evaluation Forms . . . . .	32
Statistical Procedures . . . . .	40

Chapter	Page
4. ANALYSIS AND RESULTS . . . . .	42
Analysis . . . . .	42
5. CONCLUSIONS AND RECOMMENDATIONS . . . . .	66
Recommendations . . . . .	71
REFERENCES . . . . .	73
APPENDICES . . . . .	78
A. Questionnaire To Instructors Participating in Evaluation . . . . .	79
B. Procedures For Instructors Participating in Student Evaluation Project . . . . .	81
C. Instructions To Students For Student Constructed Items . . . . .	88
D. Instructions To Students For Standardized Evaluation . . . . .	93
E. Instructor Summary Sheet . . . . .	97
F. Examples of Student Constructed Items, and Classification of 89 Student Constructed Items . . . . .	99
G. College Summary Sheet . . . . .	107
H. Instructor Report For Student Constructed Items . . . . .	109
I. Instructor Report For Kansas State Evaluation . . . . .	111
VITA . . . . .	113

## LIST OF TABLES

Table		Page
3.1	Personal Characteristics of Faculty Participating in the Study . . . . .	28
3.2	Professional Experience of Faculties . . . . .	29
3.3	Area of Highest Degree Held by Those Participating in Study . . . . .	30
3.4	Items Identified As Common Items . . . . .	37
3.5	Negatively Worded K.S.U. Evaluation Items . . . . .	39
4.1	Summary of Analysis of Variance For Eight Measures of Evaluation . . . . .	43
4.2a	Rank of Subsets For Common Item Scores . . . . .	47
4.2b	Rankings For Instructors Based on Membership in Subset For Common Items . . . . .	48
4.3	Comparison of Rankings of Instructors For Eight Measures of Evaluation . . . . .	49
4.4	Correlation Matrix of Eight Variables . . . . .	50
4.5	Correlation Matrix for Eight Measures of Evaluation With Deletion of Twelve Instructors Who Had Significantly Different Means on The Two Single Item Evaluation Forms . . . . .	53
4.6	Kendall Coefficient of Concordance For Six Measures of Evaluation . . . . .	56
4.7	Kendall Coefficient of Concordance For Five Measures of Evaluation . . . . .	57
4.8	Average Ranks of Groups For The Eight Measures of Evaluation . . . . .	61

## LIST OF FIGURES

Figure		Page
4.1	Instructor Means In Ranked Order For Common Items . . . . .	45
4.2	Combinations of Evaluation Measures For The Kendall Coefficient of Concordance . . . . .	55
4.3	Dendogram for Sixty Instructors Grouped by the MAX Procedure . . . . .	60
4.4	Rankings of Instructors Identified From Pattern Analysis for the Eight Measures of Evaluation . . . . .	62

## Chapter 1

### INTRODUCTION

Perhaps no other topic in the entire realm of higher education received more emphasis in the 1960's and early 1970's than the evaluation of college teaching. In an age of accountability, evaluation of faculties became increasingly important not only to the receivers of instruction but also to those responsible for administering institutions of higher education and to the general public, often referred to as the taxpayer.

According to each of these groups, evaluation was to serve a different purpose. Students in higher education viewed evaluation as a necessary aid in selecting classes. College and university administrators needed quantitative data to serve as a basis for decision making for salary increases, tenure, and promotion. Legislators, governing boards, and the general public saw evaluation as a method of holding public institutions of higher education accountable for expenditures of public funds (Centra, 1971:5). The increasing cost of public education in the 1970's resulted in a greater emphasis on evaluating faculty members, since faculty salaries involve large amounts of public funds.

That the consumer should be free to evaluate the product and to determine the usage of that product accordingly is a generally accepted premise. Thus much has been said and written about evaluation of teachers and their "products" by students, the "consumers." According to Johnson and reemphasized by Miller (1972:26),

It is often the individual student who knows best whether or not he is learning. It is the student who knows what is being discussed. It is the student who knows a course is stimulating him to learn more about a subject or whether it is boring him to death. It is the student who can best formulate those fundamentals and personal questions so bothering him that he cannot proceed to other academic matters. It is the student who can best evaluate when he is beginning to integrate the process of learning with the problems he continually confronts in life.

In American higher education evaluation procedures have ranged from informal student-initiated and student-directed evaluations to those developed by departmental committees and occasionally to those developed by specialists in educational measurement (Costin, Greenough and Menges, 1971:511).

The factor that is common to all evaluations of instructors is that some type of form must be completed by students of each instructor. According to Centra (1971:5), the objectives of these evaluative procedures vary from providing students with indicators of good and bad courses or teachers to providing administrators with quantitative measures of effective teaching to providing the teacher with a basis for self-improvement and professional excellence.

Faculties in general have been less than enthusiastic toward student evaluations and have viewed them with an encompassing pessimism. Such evaluations, when viewed by the faculty, tend to become an emotional issue and may consequently result in defensive measures. The faculty member may view evaluation as a threat to his job security and self-perception. Grasha (1972:1) has said that the human tendency is to resist such threats through arguments such as the following:

1. Unhealthy competition may arise among faculty members.
2. Teaching processes may become stereotyped.
3. Student inputs will not be valid since they can be misleading or inadequate.
4. No one knows what constitutes effective teaching.

Costin, Greenough and Menges (1971:511) list other frequently-cited criticisms of student evaluation of instructors:

1. Student ratings are unreliable: the ratings will be higher for an entertainer than for an instructor who gets his material across effectively.
2. A high correlation exists between ratings and expected grades: an instructor who grades strictly will receive lower ratings.
3. Students cannot judge instruction competently since long-term aspects of the course may not be evident at the time of evaluation.

The evaluation instruments used by students to evaluate their teachers have frequently been criticized. Instructors offer the criticism that the rating instruments have been developed by individuals or groups who are not highly skilled or who have not had training in the construction of such instruments. A second criticism offered by faculties is that some students apparently do not read or answer evaluation items consistently. On the other hand, students frequently criticize evaluation instruments because they do not reflect many of the qualities which students feel are vital to the learning experience.

Administrators, in the evaluation of instruction, frequently request that the results be submitted in the form of a single score. This method of reporting provides them with a method for differentiating between instructors. In practically all cases the method of differentiating between instructors is based on a score obtained from all classes taught by the instructor from whatever evaluation instrument happens to be in use at that time at that particular institution.

Too often, the student is asked to evaluate the instructor on a set of preconceived standards rather than on what the student feels is important about the instruction received. The student may be asked to add comments to some forms, but in general, he is not allowed to initiate evaluation items so that scores for the faculty member would

accurately represent the student's own perceptions of the instructional process.

### Statement of the Problem

For this study, the following aspects of faculty evaluation will be examined:

1. The use of a summated rating\* to determine differences between college instructors.

2. Student input into the evaluation process through several types of evaluation instruments.

Specifically the study will attempt to answer the following questions:

1. To what extent is it justifiable to distinguish among instructors on the basis of summated ratings?

2. What questionnaire item topics are indicated by students as basic to the ratings of all or nearly all instructors?

3. To what extent do summated ratings based on student-constructed questionnaire items correlate with those obtained from other evaluation instruments?

4. Do instructor profiles based on summated ratings from student-constructed and other rating scales provide a basis for distinguishing among instructors on the basis of

---

\*In this study, a summated rating is defined as the mean response level across questionnaire items and across students for an instructor. If a response was missing, the mean score was obtained only from those responding.

subject matter, specialization, degree level, experience, sex, and age?

### Need for the Study

The literature on the evaluation of faculty members contains many examples of research on student input into the evaluation process. A large number of these studies may be classified into three groups: studies which deal mainly with faculty concern for students as evaluators, studies analyzing the items by which the faculty member is evaluated, and studies of methodology and application. In almost all of these studies, the student is a critical factor in the evaluation process.

In the research reviewed, there appeared to be relatively little emphasis on student-originated input into the evaluation process. The primary concern of the research appears to have been on student input into an evaluation form constructed by some element of the faculty and/or administration or by an outside group. No studies were found which examined student-constructed evaluation items or forms.

In almost all the research examined, the concern focused on the validity or lack of validity of using the student as an evaluator. A large number of the studies indicate that student evaluations are used by administrators for various purposes. Again the studies fail to show or describe a methodology for use of student evaluation by

administrators. A technique for grouping faculty as a result of direct or indirect student input would provide a much-needed tool for the administrator who is concerned with intelligent decision making in the evaluation process.

## Chapter 2

### REVIEW OF THE LITERATURE

A review of the literature indicated that studies of teacher performance began around 1900. Evaluation studies after that time were sporadic until the late 1950's and early 1960's when there was a sudden new emphasis on evaluation. Recently journals have been inundated with articles on the subject of evaluation. The rating of instructors in higher education accounted for 143 entries in the Education Index for the last decade (Siebring and Schoff, 1974:150).

Examination of the literature shows relatively few books written on the subject of evaluation. In contrast, the number of journal publications and papers presented at national meetings on evaluation is overwhelming. The literature presents investigation of almost every facet of the evaluation system. Findings from many of these studies have shown no decisive results, while the findings from others have been contradictory.

#### Early History

Early in the literature on the preparation of teachers, various methods were tried in order to ascertain the perfect teacher (Medley, Mitzel, [Gage, 1963:248]). A widely held belief asserted that a trained supervisor or

instructional expert could measure the effectiveness of a teacher by watching him teach.

The earliest studies of teacher characteristics were investigations by W. F. Book in 1906. The study by Book concerned high school students. He (Boyce, 1920:325) concluded that:

1. Pupils demand constant direction and encouragement.
2. The common virtues appeal to high school students.
3. The teacher's ability to understand her pupils is very important.
4. Pupils like teachers of the same characteristics as themselves.
5. The sex of the instructor is not a factor.

Remmers, noting the emphasis on research in teaching, developed the Purdue Rating Scale in 1926 (Gage, 1973: 367). In its revised form for student evaluation of college faculty, this instrument is based on the thesis that student evaluation is a useful, convenient, reliable, and valid aid to the teacher's self-improvement.

An era during which considerable money was spent to improve education began in 1957 with the launching of Sputnik. This emphasis in education resulted in the pouring of many millions of dollars into the education industry. More dollars brought more students, more faculty, and more different pressures into higher education. Soon anguished

cries over expenditures or citizen disgust with the appearance of students brought efforts for accountability.

Evaluation of faculty members was emphasized as a part of this accountability process. Accounting for teacher effectiveness in the classroom has brought a new wave of research into this particular area (Miller, 1972:5).

These concerns of the last two decades have produced a constant surge of literature on all facets of the problem of evaluation. Therefore, the review of literature for this study will include only the following areas which are relative to this research:

1. Uses of Student Evaluators of Instruction.
2. Criticisms of Methods and Use of Student Evaluation.
3. Student Input into Instructor Evaluation.
4. Summated Ratings and
5. Standardized Rating Scales.

#### Uses of Student Evaluators of Instruction

One college president described faculty evaluation to the Southern Regional Educational Board as "being frightening because those in education have not yet devised a system which is clearly understood and accepted" (Maiher, 1975:5).

The question of what constitutes good teaching is not one that is easily resolved. Conscientious administrators are continually troubled by the problem of

"recognizing and rewarding good teaching" (Bryant, 1967:327). Assuming that certain criteria could be defined to characterize good teaching, we would still face the problem of the method of use and the question of who should be the judge of effective teaching: the students, faculty or administrators (Walker, 1974:18).

Educators who have previously ignored the issue of evaluation have been forced to recognize its importance; for example many legislatures have recently passed bills mandating distinct evaluation procedures (Cohen, 1972:33a). Cohen (1972:33c) states four purposes for evaluation: "to make judgements about faculty, to award merit pay, to provide a basis for establishing tenure or continuing contracts and to provide evidence of faculty competency." It is apparent from this list that improvement of instruction does not share a top priority for conducting faculty evaluation.

Because of the decreasing mobility of faculty, evaluation is receiving greater emphasis. Previously, those denied tenure would move on to positions in other colleges with little difficulty. However, since faculty members are now remaining at an institution for the number of years requisite to tenure, administrators are finding that they must establish a basis for granting or denying tenure (Maiher, 1975:1).

A justification for using student evaluation of instruction is that it provides students a voice in their

educational experience and that it gives them an opportunity to participate constructively in the educational organization. Student evaluation serves as a channel of communication for substantive information regarding educational policies (Kent, 1966:399).

Student evaluation of instruction is often described as an explosive issue. Epstein (1974:32) describes the three groups who are affected most directly by evaluation in this way: students endorse it, administrators favor it for accountability, and faculty rise in a call to arms if it is even suggested.

#### Criticisms of Methods & Use of Student Evaluation

Faculty members have long contended that students do not know what good teaching is and therefore cannot make reliable judgements. Others argue that because the student is immature, his objectivity will be influenced by extraneous matters (Walker, 1974:18).

An anonymous statement quoted by Laura Kent (1966:400) presents a representative view of the arguments generally given against student evaluation and its use.

To the question "Can the student evaluate his teacher?" the simple answer is, of course, no. Partly because he has not the ability, partly because he has not a real opportunity, partly because no scheme for recording his evaluation can be both fair to him and intelligible to anyone else. Going further, I may add that he has no right to even try and that encouraging him to do so is simply to aggravate his already considerable power of mischief.

The concern for the practice of using student evaluations of the quality of instruction was emphasized in a recent article by Rodin and Rodin (1972:1166). Their study indicated that students rate high the instructors from which they learn least.

Another argument frequently given against evaluation is that the evaluative instruments are ineffective. Kent (1966:402), after reacting to the Muscatine Report, suggested that, because of bias and unreliability, evaluation instruments may reflect aspects of teacher performance that are not central to the basic purposes of education. These instruments are also described as being subject to human limitations such as faculty memory, inaccuracy of observations, and selective perception.

Greenwood and others (1973:596) found that most correctly-used evaluation instruments require the student to specify the degree to which a statement characterizes an instructor. He points out that instruments such as the Purdue Instructor Performance Indicator usually use a forced choice format which requires the student to assess global characteristics of the faculty member. This method is limited in terms of student response sets and instructor halo effect. Thus authors suggest the use of an instrument which (1) would be empirically derived but which reflects a broad conception of college instruction and (2) would focus on specific teaching behavior and permit students to rate only those items which they consider to be relevant.

Another method of obtaining student ratings was described by Frey (1973:147-150). He recommended that student ratings be obtained through a questionnaire which would be at least partially designed by the instructor. Frey defined the following specific requirements for the questionnaire:

- (1) The instrument should contain 20 items or less.
- (2) Items should be sufficiently general so as to apply to all subject areas.
- (3) Topics in which students have limited expertise should be avoided.

#### Student Input into Instructor Evaluation

Remmers (Gage, 1963:367) stated that student ratings of teachers is one means of enabling the teacher to see himself through the eyes of his students. His quotation from "To a Louse" by Robert Burns describes the need for feedback from others this way:

Oh wod some power the giftie gie us  
To see ourselves as ithers see us!  
It wod frae mony a blunder free us  
And foolish notion

Remmers reported that numerous studies using the Purdue Rating Scale for instruction have established student ratings of teachers to be reliable and valid (Gage, 1963:367-68). Costin, Greenough, and Menges (1971:530) reported that numerous investigations found acceptable stability and internal consistency of student ratings, and

that responses were not biased by experience unrelated to the course. Their findings suggested that the criteria used by students in their ratings of instructors dealt more with the quality of the material presented than with the entertainment value of the course. The attributes of clarity of presentation, preparation, and intellectual stimulation were usually cited by students in describing their best instructors. Costin (1971:530) observed that the correlation between course rating and grade received was small. He noted that other studies indicated a relationship between grades and course rating tended to result from greater student interest than the so called reward effect. Kent (1966:403) indicated that most studies found that the grade a student expected to receive in a course was not related to the rating of the instructor. The study by Remmers (Gage, 1963:367) found little if any relationship between these two factors. Kent also noted that in a study conducted at the University of Michigan students tend to agree on the ratings of very good or very bad teachers but show less agreement on those in the middle. Costin and Kent disagree on the questions of required classes being rated more severely than those classes taken only by majors, and of teachers with higher academic ranks receiving higher ratings than their less experienced colleagues.

Findings from a study by Doyle and Whitely (1974:259-274) showed evidence that student characteristics such as years in school, sex, and ability do not influence

student opinion of overall instructor characteristics, but do influence ratings about specific teacher characteristics. The authors concluded that mean ratings of the general teaching ability or overall effectiveness are valid criteria for teaching effectiveness.

Another segment of the evaluation process deals with the qualities students consider to be of greatest importance in teaching a course. A study conducted at the University of California in 1969 determined that the following eight teacher traits were most important.

1. A dynamic and energetic personality.
2. Clear explanations.
3. Interesting style of presentation.
4. Enjoyment of teaching.
5. Interest in students.
6. Friendliness toward students.
7. Encouragement of class discussion.
8. Discussion of points of view other than his own (Miller, 1972:24).

In a study conducted at the State University of New York (Musella and Rusch, 1968:138), students were asked to identify the three qualities they considered to be most important for teachers in the following course areas:

1. Physical and biological sciences.
2. Social sciences.
3. Arts and literature.

For the physical and biological sciences students identified:

1. Ability to explain clearly.
2. Systematic organization of subject matter.
3. Expert knowledge of subject.

The three qualities for the subject areas of the arts, literature, and the social sciences were the same:

1. Ability to encourage thought.
2. Enthusiastic attitude toward subject.
3. Expert knowledge of subject.

Students at Philander Smith College (Pogue 1967: 133) cited knowledge of the subject and careful evaluation (testing) as qualities that best describe the ideal professor.

#### Summated Ratings

Kenneth Eble (1974:455) points out the emphasis that is given to the mechanical aspects of evaluation. He considers it a danger for student ratings to be used as single number indices of faculty competence--that a professor who receives a rating of 3.7 will be compared with another professor who receives a 3.9 rating. He calls for specific feedback on various characteristics of teacher performance. He further states that "reducing multiple information to a quantitative single score does not outweigh the value of the information that good evaluation questionnaires can provide."

Nelson Cossart (1974:11) states that the "real advantage of applying the method of summated ratings to an evaluation form is that an individual can assess his position with respect to the total group. It allows a comparative evaluation where known standards do not exist." These writers point to the disadvantage of a system that compares one person's scores with the scores of an average group. Half the group must fall above and half below the average. While students may provide high ratings of many instructors, some instructors would nevertheless occupy a low relative position.

#### Standardized Rating Scales

Standardized means of evaluating instructors first appeared in the 1920's. One of the oldest scales in existence, the Purdue Rating Scale for Instruction, was first developed in 1926. This instrument has been subject to extensive and intensive research through its development. Miller (1972:9) states that the conclusions drawn from the research show "a third of a century of use . . . by many teachers and a very considerable amount of experimental research . . . have demonstrated that student evaluation is a useful, convenient, reliable, and valid means of self-supervision and self-improvement for the teacher."

Hoyt (1973b:73-76) in an intensive study of measures of instructional effectiveness defined a set of

eight objectives of undergraduate instruction at Kansas State University.

1. Gaining factual knowledge.
2. Learning fundamental principles and theories.
3. Applying principles to problem solving.
4. Understanding oneself.
5. Learning professional attitudes or behaviors.
6. Developing communication skills.
7. Utilizing implications of the course for professional development.
8. Gaining a liberal education.

From his eight objectives, from analyses of evaluation instruments used by other researchers, and from recommendations of student and faculty evolved the first 58 items of Kansas State University Evaluation. Results of use of the instrument have been favorable for evaluating instructional effectiveness. Hoyt views the Kansas State University evaluation as a means of helping instructors improve their teaching rather than as a tool for the administration.

A relatively new instrument, the Student Instructional Report, was developed in 1972 by John Centra of the Educational Testing Service from extensive research on student evaluation of teaching and faculty development (Beck, 1975). This instrument focuses on organization and structure of course, teaching techniques and student teacher rapport. In addition, the instrument also included questions on demographic data of the student.

Beck emphasized that the major strength of the instrument is that it is a part of a continuous research program and that the principle purpose of the Student Instructional Report is to improve instruction.

Another recent development in standardized evaluation instruments is the Purdue Cafeteria System developed by the staff of the Measurement and Research Center at Purdue University (Siebert, 1975). The purpose of this instrument was similar to that of the Kansas State Evaluation Instrument. The chief difference between the two instruments is that the Purdue system allows the instructor to select 40 items from a computer based "menu" of 200 items that he feels best describe his instructional situation.

Siebert notes that the instrument also has five mandatory core items which include student motivation, explanation of material, stimulating assignments, comparison of course with other courses, and comparison of instructor with other instructors. Siebert emphasized the fact that the Purdue Cafeteria System provides diagnostic information which can be used for the improvement of instruction.

In summary, Eble (1974:456) cautions faculty members to end fears about student evaluation.

"Its dangers are not that it will do too much harm but that it will do too little good. In the present apathetic student climate, I have more to fear that the student will give it up than the faculty will be harmed."

## Summary

Research studies reviewed in the literature indicated that the use of students as evaluators of instruction was justified on the basis that the student should be given an opportunity to evaluate the instructional process. Faculties in general have been dissatisfied with students as evaluators and have offered many arguments against the use of student instructional evaluation. Lack of knowledge and ability to effectively evaluate instruction was an argument frequently given against student evaluation.

The use of summated ratings was criticized because a single score often obviates individual items for assessment of the instructor. Comparisons among instructors based on summated scores refer to an average which would cause an instructor, who scored at a below average level, to be placed at a disadvantage even though the evaluation may have been comparatively high.

Standardized forms for use in the evaluation of instruction by students have been in existence since the mid-twenties when the Purdue Rating Scale for instruction was developed. Since that time a number of standardized scales have come into existence. Of each of the scales examined, the stated purpose was the improvement of instruction.

## Chapter 3

### RESEARCH DESIGN AND PROCEDURE

One means of responding to the accountability syndrome of the 1970's was the widespread use of formal student appraisal for the evaluation of instruction. The means of obtaining student input was the object of much debate in faculty evaluation systems. Most instruments used for student evaluation of instruction have required the student to respond to a set of pre-conceived statements or items. Another less frequently used method of obtaining student input required the student to react to the evaluation process by indicating the elements of instruction that were important to his own learning process.

In this study two methods of evaluation were examined using three different evaluation instruments:

1. A form on which students constructed their own evaluation items and rated the instructor on each item based on a four-choice scale.
2. A standardized evaluation form developed by Kansas State University.
3. A single-item evaluation form which requested the student to provide an overall rating of the instructor.

The procedures utilized in this study were developed in three phases. Phase one involved direct student input with respect to the following goals:

1. To classify student-constructed items into categories representing extremely similar item content for a given instructor.

2. To identify item categories that were prevalent for each instructor.

3. To identify from among the item categories prevalent for each instructor those categories that would apply to all or nearly all instructors. These categories are referred to hereafter as "common items" (for all instructors).

4. To identify for each instructor the prevalent item categories not classified as common items. These item categories are referred to hereafter as specific items (for each instructor).

5. To identify for each instructor the individual student-constructed items or item categories not classified as common or specific. These are referred to hereafter as idiosyncratic items (for each instructor).

6. To determine summated ratings for each instructor based on the three item sets: common items described in 4 above, specific items described in 5, and idiosyncratic items described in 6.

Phase two involved indirect student input into faculty evaluation through the use of the standardized evaluation form developed by Kansas State University (Hoyt, 1973a). This form is composed of three parts; however, only part one, which contains 58 true-false items concerning the instructor and the teaching process, was used for this study. Items from this form were used to obtain summated ratings for this phase of the study.

The third phase of the study was based on faculty evaluation through the use of a single-item evaluation form. On this form the student rated the instructor on a one-to-ten scale with one being the lowest rating and ten being the highest rating.

#### Population and Sample

Students and faculty members from three colleges in southwestern Virginia participated in the study. The colleges were Radford College, Virginia Polytechnic Institute and State University, and New River Community College. These three colleges offer a wide variety of programs and degrees.

Radford College is a comprehensive state college located in Radford, Virginia, with programs which lead to Bachelor's and Master's degrees. Curricula in liberal arts, fine arts, and education are emphasized. In the fall of 1974 the college had 231 faculty members and 4300 students.

Virginia Polytechnic Institute and State University

is a comprehensive state university with 18,000 students, located 40 miles southwest of Roanoke in Blacksburg, Virginia. The university offers a large variety of graduate and undergraduate programs in engineering, agriculture, arts and sciences, education, architecture and home economics.

New River Community College, with an enrollment of 1900 students and approximately 75 faculty members, is an urban-rural community college which serves a five-county area in the New River Valley of southwest Virginia. The college offers transfer curricula and a wide variety of occupational-technical programs leading to certificates, diplomas, and associate degrees.

Permission to conduct the study at the three colleges was obtained from chief administrative officers. Winter quarter schedules for 1975 were consulted to identify instructors who taught two sections of the same or two similar courses. These courses were limited to freshman and sophomore level classes, although these sections may have included upper classmen at the two senior institutions. Because the number of instructors having two similar classes varied widely among the three institutions, it was decided to use a sample of twenty instructors from each college who met the class requirements. The instructors from each institution were selected by use of the random

number table. Faculty members selected were contacted and asked if they would participate in the study. Most faculty members were willing to conduct the study in their classes.

Some faculty members originally selected as part of the sample population did not participate. At Radford College two faculty members requested not to participate because of lack of class time near the end of the quarter. Two instructors from Virginia Polytechnic Institute and State University would not participate in the study, two did not return the evaluation materials by the stated deadline, and two faculty members did not conduct the study according to directions. At New River Community College two faculty members did not return evaluation materials within the time period. In all cases, replacements for those faculty not participating or failing to return evaluation materials were selected using the same criteria and method as the original participants.

The student subjects for this study were those enrolled in the two similar classes of the selected instructors. Students present in the designated classes on the day on which the evaluation forms were distributed were invited to participate in the study.

#### Profile of Faculty Participating in the Study

Each faculty member participating in the study was requested to complete a short questionnaire of personal characteristics and professional experience (Appendix A).

Personal items included age and sex. Professional experience items included the highest earned degree, degree field, and amount of teaching experience.

The mean age of participants ranged from 38 at Radford to 31 at Virginia Polytechnic Institute and State University. Because some graduate teaching assistants at the University participated in the study, the average age for this group tended to be lower. Faculty members were evenly divided between male and female at the community college, while more male faculty members participated in the study from the two senior colleges (Table 3.1).

The instructors having the greatest amount of teaching experience were those at Radford College. Virginia Polytechnic Institute and State University faculty tended to be younger and, therefore, to have fewer years of teaching experience. The senior colleges teachers also had the largest number of doctorates. The majority of faculty participating in the evaluation from the community college held master's degrees (Table 3.2). Areas in which the highest earned degree was held covered a wide variety of academic and technical areas (Table 3.3).

### Procedures

From the classes taught by each instructor, the writer randomly selected one class in which the instructor would administer the Student Constructed Items Evaluation Form plus the single item evaluation form. In the other

Table 3.1

Personal Characteristics of Faculty  
Participating in the Study

Characteristic	N.R.C.C.		Radford		VPI & SU	
	N	Mean	N	Mean	N	Mean
Age	20	36	20	38	20	31
Sex:						
Male	10		16		14	
Female	10		4		6	

Table 3.2

## Professional Experience of Faculties

Experience & Degree	N.R.C.C.		Radford		VPI & SU	
	N	Mean	N	Mean	N	Mean
Years employed by college	20	6.3	20	7.2	20	3.4
Number of years teaching experience	20	8.8	20	10.2	20	6.3
Highest earned degree						
Diploma	1					
Bachelors	5				2	
Masters	14		6		8	
Doctorate			14		10	

Table 3.3

Area of Highest Degree Held by  
Those Participating in Study

Area	NRCC	Radford	VPI & SU
Art		1	1
Auto Mechanics	1		
Biology		3	
Business Education	3		
Dance		1	
Economics		1	
Education	2	1	1
Education Administration	1		
English	2	3	3
Food & Nutrition			1
Geography		1	1
Geology		2	1
Health & P.E.	1		
History	2	2	1
Law Enforcement	1		
Mathematics	2	2	2
Music		2	2
Philosophy		1	1
Political Science			1
Psychology	1		
Social Work	1		
Sociology			1
Statistics			2
Voc.-Tech. Education	2		

class, the instructor was requested to give the Kansas State Evaluation Form and the single item evaluation form. The instructors were given a written set of instructions for administering evaluations in each class. In addition to the written instructions, a cassette tape illustrating examples of evaluative items was made available to the students writing the student constructed evaluation items. (Copies of instructions for faculty and the text of the cassette tape are found in Appendix B). Evaluation materials required to administer the instrument were distributed to instructors during the Winter Quarter, 1975.

In the class which was designated as the student-constructed items class, the students were instructed to write as many evaluation items as they felt necessary to adequately evaluate the instruction they had received in that class and then to evaluate the instructor on those items. Items were to be written so that they could be evaluated on a four category graphic scale: always, usually, sometimes, and never (Appendix C).

After completing the written evaluation, the students were asked to evaluate the instructor using a single item evaluation form. This form specified that the instructor should be evaluated in that course in comparison with other instructors that the student had been associated with in college. Other criteria to be considered in the evaluation included: effectiveness in presentation of subject matter, knowledge of subject matter, fairness in

grading, commitment, enthusiasm, and availability to students. A scale of 1 to 10 was used with one being the lowest rating and ten being the highest rating (Appendix C).

In the other class designated for evaluation, the instructor was requested to administer the Kansas State University evaluation form. Students were asked to evaluate the instructor as accurately as possible on each of the 58 items found in part one using the designated true-false scale (Appendix D).

When the Kansas State University evaluation was completed, the students were then given the single-item evaluation form and asked to evaluate the instructor as previously described.

#### Processing of Evaluation Forms

The evaluation forms were processed as follows:

1. Student-Constructed Items Form. The four category scale--always, usually, sometimes and never--used for rating instructors was assigned numerical values from four to one with "always" equal to four and "never" equal to one. Negatively worded items were reversed to provide consistency with the scale. In the classification of items, if a student wrote two items that would fit into one classification and gave the instructor the same rating on each item, the rating was recorded only once. If the student wrote two or more items which fit into one classification and gave the

instructor different ratings on the items, a mean score was recorded.

A summary sheet was prepared to record individual student responses for each instructor. This sheet provided space for recording item categories and the numerical values of each category for each student in the class (Appendix E).

The writer randomly selected a set of student-constructed forms for one instructor to serve as a starting point for analysis. Items for the first student were recorded on the summary sheet along with the numerical rating for each item. Items written by the remaining students in the class were then matched into the categories of items previously listed, if appropriate. If the item did not fit into the item classification previously listed, it was added as a separate category.

The original plan called for the use of a panel of students to classify all student-constructed items into categories. However, once the evaluation forms were received, an inspection of the student-generated items revealed that for the most part they could be readily classified into an item category by the writer. For example, the following items, directly quoted from student evaluation forms, were classified into the category of "Interesting Presentation":

1. Instructor makes lectures interesting.
2. Adds humor to brighten class.

3. Class never boring.
4. Keeps class interested.

Items that were classified in the category of "Student-Oriented" or "Relates To Students" included:

1. Gets along well with students.
2. The teacher seems interested in the students.
3. Interested in you as a person.
4. Instructor was concerned with students rather than self.

The categories of grading fairness and test fairness are closely related as the lists compared below indicate. Items for these categories were usually classified by the use of the words tests or grades.

<u>Grading Fairness</u>	<u>Testing Fairness</u>
1. Fairness in grading.	1. Thorough tests.
2. Teacher keeps students informed on grades.	2. Teacher provides ample time for tests.
3. Teacher lets students know exactly what is expected of him.	3. Well prepared tests.
4. Grades equally among students.	4. Prepares students well in advance for tests.
5. Gives you the opportunity to help your grade.	

The fact that items generated by students were so easily classifiable may have been partly due to the fact all students had previously participated in evaluation of instruction. All three colleges represented in the study

had administered some type of evaluation during the preceding quarter. Other examples of student constructed items and the 89 classifications of items appear in Appendix F.

Student-constructed items such as those listed below were eliminated from the study because they indicated factors that did not relate to teachers or were beyond the control of the teacher.

1. Has the teacher too many students?
2. Class evaluation is an important part of this college.
3. The class, through no fault of the teacher, is lazy.
4. Texts are usable and applicable if required to be purchased.

Some students apparently did not view the exercise of composing items as a completely serious process. Such items as "sings to us," for a non-music teacher, and an item addressed to appearance, "mighty easy on the eyes," expressed student observation of non-instructional qualities.

After categorizing or eliminating all items as described above, a summary sheet was prepared which listed all item classifications used by the students to evaluate the instructors in each college (Appendix G). The sum of the item ratings and the number of students responding to that item were recorded for each item for each instructor.

Three item sets were identified from these summaries:

Common items. Those item categories common to at least 90 percent of all instructors and on which at least 25 percent of the students in each class evaluated the instructor were included. Only five of the 89 item categories met the criteria for the common items (Table 3.4).

Specific items. Item categories which include evaluation items not defined as common, yet on which 33 percent of the class evaluated the instructor were included. The number of specific items varied from instructor to instructor. In some cases an instructor may have had only one or perhaps no specific items.

Idiosyncratic items. Those items which did not meet the criteria for one of the above classifications were included.

In addition to these item sets, the common items and specific items were combined to produce a fourth item set (common/specific items). A fifth item set was obtained by combining common, specific, and idiosyncratic item categories (grand total).

Summated scores were computed for each student for each of the five item subsets described above. For each student a summated score for each subset was obtained by adding together the ratings for each item produced by the

Table 3.4

## Items Identified As Common Items

- 
- 
1. Knowledge of material
  2. Interesting Presentation
  3. Teaching Method - Ability to  
Explain Material
  4. Grading Fairness
  5. Orientation Toward or Relationships  
with Students
-

student within that subset and dividing by the number of these items.

A report covering specific items, common items, and the frequencies of rating scale responses to each such item was prepared for instructors participating in the study (Appendix H).

2. Kansas State University Evaluation Form. Only the first 58 items of the Kansas State University Evaluation Form were used for this study. Five of these items (45, 46, 49, 56, and 57) were omitted from the analysis because they did not relate directly to the evaluation of instruction.

Nineteen items were negatively worded (Table 3.5). To provide a consistent measure, the scoring for these items was reversed. For scoring purposes, items marked true were assigned a value of one and items marked false were assigned a value of two. A mean score was obtained for each of the individual items for each instructor. The item means were not used in the data analysis, but were used to report results of the evaluation to the instructor. An example of an individual report is shown in Appendix I.

A summated score for each student for each instructor was obtained by using the numerical values for the true-false items and dividing by the total number of responses. Again, omissions were not counted in the computation of the summated score. The values for the non-usable questions were also omitted from the computation of the summated score.

Table 3.5

## NEGATIVELY WORDED K.S.U. EVALUATION ITEMS

- 
- 
4. The instructor seemed to lack energy.
  9. He was often incoherent and/or vague in what he was saying.
  11. He generally spoke too rapidly.
  13. On several occasions, he seemed unprepared for class.
  17. His presentations were dry and dull.
  23. He lectured in a low monotone.
  25. He failed to state clearly the course requirements and deadlines.
  29. He lectured in a rambling fashion.
  32. He became angry or sarcastic when corrected or challenged by a student.
  33. He failed to differentiate between significant and non-significant material.
  35. He repeated material to the point of monotony.
  36. He displayed favoritism.
  51. Examinations stressed memorization of information for which later recall seems unreasonable.
  55. Examination questions were frequently too detailed or picky.
  58. There were too many topics to understand any of them well.
- 

Items 45, 46, and 49 were also negatively worded but these items were omitted from the study.

3. Single-Item Evaluation Form. The second type of form used in this evaluation was the ten-point rating scale. This scale was given along with both the student-constructed items form and the Kansas State Evaluation Form. The rating identified by each student became his score on this instrument. An omission of this score was treated simply as an omitted score and did not disqualify the student's responses on the other evaluation form.

#### Statistical Procedures

Summated scores from individual students were used as the bases for all analyses. Means across students of the summated scores for each instructor were used to test differences between instructors. Means of summated scores from the two evaluation forms plus the student-constructed form and subsets of the student-constructed form provided eight measures of evaluation for the sixty instructors. These measures were based upon student responses in the following areas:

1. All student-constructed items.
2. Common items.
3. Specific items.
4. Idiosyncratic items.
5. Common/Specific items.
6. Single item evaluation for class with student-constructed items.
7. Kansas State University Evaluation.

8. Single-item evaluation for the class with Kansas State University evaluation.

A one way, univariate analysis of variance was used to analyze means of student summated scores for each instructor on each of the eight evaluation measures. Duncan's New Multiple Range Test (Duncan, 1955) was used to isolate groups without significant mean differences in each of the eight evaluation measures.

Using the ranks of each instructor's means on each of the eight measures listed above, Kendall's coefficient of concordance was computed to determine the degree of agreement between the eight measures for the sixty instructors. These ranks were also used to obtain correlations among the eight measures.

A pattern analysis using Johnson's MAX procedure (Baker, 1972) was made to determine yet another aspect of groupings of instructors. The rankings were again used to determine whether rank profiles on the eight measures could be related to personal and professional characteristics of instructors.

## Chapter 4

### ANALYSIS AND RESULTS

This chapter contains an analysis of the data for sixty instructors from three different types of colleges for three different types of evaluation instruments.

#### Analysis

One way analyses of variance for unequal group sizes were used to determine whether there were significant differences among the means of the student response measures of the sixty instructors on the different measures of evaluation. A separate analysis was performed for each of the eight measures of evaluation. These analyses were evaluated at the .01 level of significance. There were significant differences between the means for instructors on all of the measures of evaluation. Analyses were significant beyond the .01 level of probability for each of the eight measures of evaluation (Table 4.1).

These significant overall analyses of variance were further investigated by a multiple comparison procedure which tested all pair-wise comparisons among means, the Duncan New Multiple Range Test (BMC 07V, Dixon, 1973:677-692) for unequal sample sizes. (This is an extension of the multiple range test described by Kramer 1956:307-310).

Table 4.1

## Summary of Analysis of Variance For Eight Measures of Evaluation

Measure	Source of Variation	df.	ss	ms	F*
SCI Common	Between	59	7,427.0	125.9	3.74
	Within	1134	38,214.7	33.7	
SCI Specific	Between	58	10,784.3	185.9	4.66
	Within	893	35,560.3	39.8	
SCI Idiosyncratic	Between	59	8,113.7	137.5	3.19
	Within	1161	50,007.8	43.0	
SCI Common & Specific	Between	59	8,940.7	151.5	5.69
	Within	1204	32,049.4	26.6	
SCI Overall	Between	59	7,412.9	125.6	4.79
	Within	1246	32,623.3	36.1	
Single Item Evaluation With SCI	Between	59	843.3	14.3	6.39
	Within	1216	2,719.1	2.2	
Single Item Evaluation With KSU	Between	59	1,082.8	18.3	8.54
	Within	1255	2,696.0	2.1	
KSU	Between	59	7.04	.1194	8.40
	Within	1323	18.8	.0142	

\*All were significant beyond the .01 level.

Each of the eight analyses produced similar results at the .05 level of significance. No subsets of instructors whose means were not significantly different from each other failed to share members with another such subset. An example of the overlap of subsets is shown for the analysis of the SCI Common Items (Figure 4.1).

The Multiple Range Test identifies subsets of instructors for which means of ratings by students were not significantly different. The homogeneous subsets were used in an exploratory approach to assigning fair and meaningful comparative measures across instructors. The theory for this approach assumed that differences between instructors could be determined on the basis of the ranks of the homogeneous subsets in which the instructor appeared. Subsets generated from the analysis of common items will be used to explain the procedure.

Eleven homogeneous subsets of instructors were identified from the common items. These subsets were ordered from lowest to highest on the basis of ranks of the means of the instructors making up the subset. The subset containing the lowest ranks was assigned a rank of one and the subset containing the highest ranks was given the highest number, in this case, eleven (Table 4.2a). If an instructor appeared in only the lowest subset, that instructor was assigned a rank-score of one. If the instructor appeared in subsets one and two, he was given a rank-score of 1.5 for the average of rank one and rank two. This

Instructor Rank from Mean Rating	Mean	Number of Students
1	26.000	11
2	28.038	26
3	28.192	26
4	30.133	15
5	30.163	43
6	30.538	13
7	30.889	9
8	31.000	10
9	31.091	11
10	31.737	19
11	32.333	3
12	32.700	20
13	32.700	20
14	33.067	15
15	33.111	9
16	33.200	10
17	33.364	11
18	33.438	16
19	33.462	26
20	33.500	28
21	33.526	19
22	33.533	15
23	33.625	16
24	33.722	18
25	33.755	53
26	33.846	39
27	33.941	17
28	34.188	16
29	34.214	14
30	34.333	6
31	34.391	23
32	34.462	15
33	34.476	21
34	34.577	26
35	34.625	16
36	35.095	21
37	35.321	28
38	35.370	46
39	35.500	12
40	35.545	33
41	35.571	7
42	35.600	10
43	35.600	20
44	35.842	19
45	35.909	11
46	35.919	57
47	36.000	18
48	36.342	38
49	36.400	20
50	36.542	24
51	36.643	14
52	36.700	10
53	36.750	20
54	37.000	11
55	37.357	14
56	37.545	11
57	37.643	14
58	37.694	49
59	37.750	20
60	37.765	34

Each bracket indicates a homogeneous subset.

Figure 4.1

Instructor Means In Ranked Order  
For Common Items

procedure was continued until each instructor had been given a rank-score equivalent to the mean of the ranks of the subsets to which he belonged.

Using this procedure made it possible to identify those instructors receiving very high or very low ratings. A large number of instructors still clustered in the middle ranks (Table 4.2b).

This procedure could be considered as a possible alternative for the administrator who must make decisions based on student evaluations. According to the example described above, the most favorable decision would be made for the instructor receiving the highest rank-score, the least favorable for the few receiving the lowest rank-scores. This process could lead to a mathematical formula based on the amount of money available for raises. The amount of salary increase per instructor could then be calculated according to the number of instructors and the percentage of instructors receiving each rank-score.

Comparisons of the eight measures of evaluation were made by computing correlation coefficients between the ranks of the instructors on each measure of evaluation. (Ranks for each instructor on each evaluation measure are found in Table 4.3.) Correlation coefficients between the measures of evaluation are displayed in Table 4.4. It was expected that high correlations would occur between the subsets of the rankings for student-constructed items and the Grand

Table 4.2a

Rank of Subsets For  
Common Item Scores

Rank Assigned to Subset	No. Instructors in Subset	Range of Instructor Ranks For Subsets
1	11	1-11
2	14	2-15
3	28	3-30
4	38	4-41
5	42	4-45
6	47	6-52
7	48	7-54
8	50	7-56
9	47	11-57
10	49	11-59
11	50	11-60

Table 4.2b

Rankings For Instructors Based on  
Membership in Subset  
For Common Items

Instructor Ranks from Mean Ratings	Number of Instructors	Rank-Score (Mean Sub-Group Membership Rank)
1	1	1.0
2	1	1.5
3	1	2.0
4-5	2	3.0
6	1	3.5
7-10	4	4.5
11-41	31	7.5
42-45	4	8.0
46-52	7	8.5
53-54	2	9.0
55-56	2	9.5
57-58	2	10.0
59	1	10.5
60	1	11.0

Rank of Means is shown in Figure 4.1

Table 4.3

Comparison of Rankings of Instructors  
For Eight Measures of Evaluation

Instructor #	Evaluation Measure							
	KSU	KSU-10	SCI-10	Comm.	Comm/Spec.	Specific Total	Grand Total	Idio-Syncratic
1	13	16	6	39	27	37	11	9
2	41	28	49	50	55	53	51	35
3	19	29	32	13	18	30	6	5
4	43	42	40	43	49	54	35	16
5	16	19	13	52	21	12	17	32
6	21	54	28	26	28	25	33	30
7	37	32	35	53	53	50	49	29
8	1	2	3	2	1	1	2	1
9	22	20	16	4	11	22	9	10
10	25	17	10	44	19	8	23	41
11	28	11	34	48	44	41	43	26
12	14	36	29	9	8	17	12	19
13	33	34	39	27	26	16	37	31
14	23	26	24	19	22	29	30	38
15	7	3	19	34	32	27	24	12
16	32	39	41	57	47	15	44	39
17	40	30	30	40	42	44	40	45
18	47	31	20	21	20	28	29	37
19	50	27	38	42	25	21	28	15
20	30	35	45	36	38	49	47	48
21	48	47	50	60	52	47	60	59
22	36	33	21	25	30	24	25	20
23	52	46	43	58	56	55	53	52
24	44	56	46	38	43	40	48	46
25	3	7	12	24	6	9	14	25
26	31	9	9	12	13	20	15	21
27	27	41	33	33	23	26	34	33
28	9	8	15	28	33	33	19	13
29	8	5	4	5	5	5	5	7
30	17	15	22	35	14	13	21	28
31	11	14	11	30	39	42	18	22
32	10	18	14	10	7	10	8	14
33	20	22	48	41	29	18	38	55
34	15	13	5	8	4	3	4	11
35	26	10	52	45	51	43	54	44
36	4	4	1	1	2	2	1	2
37	12	6	26	23	12	11	7	6
38	54	58	37	31	40	39	50	58
39	29	21	17	20	24	34	22	18
40	2	1	2	3	3	4	3	3
41	57	57	55	49	48	48	42	27
42	6	12	18	14	17	19	10	24
43	35	43	25	6	10	14	26	49
44	24	24	51	46	46	56	46	40
45	38	38	60	22	31	35	41	56
46	34	52	23	47	45	46	13	8
47	55	51	31	17	34	32	32	17
48	56	44	53	55	60	60	56	51
49	42	25	27	39	35	38	45	47
50	18	53	58	54	57	52	59	54
51	5	23	54	16	16	7	31	36
52	53	48	42	37	50	59	39	42
53	39	37	36	18	15	6	27	43
54	60	59	56	32	41	45	52	50
55	45	45	8	7	9	31	16	23
56	46	40	44	11	37	23	36	34
57	59	60	57	56	59	58	55	53
58	49	49	7	15	36	36	20	4
59	51	50	47	51	54	51	58	60
60	58	55	59	59	58	57	57	57

Table 4.4  
Correlation Matrix of Eight Variables

Evaluation Measures	1	2	3	4	5	6	7	8
1-KSU	-							
2-KSU-10	.8032	-						
3-SCI-10	.5765	.6245	-					
4-Common	.4685	.4067	.6308	-				
5-Common & Specific	.6522	.6008	.7317	.8466	-			
6-Specific	.6487	.5906	.6198	.6850	.8978	-		
7-Grand Total	.6876	.6381	.8595	.7708	.8639	.7426	-	
8-Idiosyncratic	.5488	.5573	.7392	.5895	.5872	.4880	.8493	-

Total for these items. (SCI-Grand Total - Common Items,  $r = .77$ ; SCI-Grand Total - Specific Items,  $r = .74$ ; SCI-Grand Total - Idiosyncratic Items,  $r = .85$ ; SCI-Grand Total - Common/Specific Items,  $r = .86$ ). It was also expected that rankings for the Common/Specific Items would correlate highly with the Common Items ( $r = .85$ ), and the Specific Items ( $r = .90$ ) since these are subsets of the Common/Specific Items.

High agreement occurred between the KSU evaluation and the single-item evaluation which was given in the same class ( $r = .80$ ). A similarly high correlation was also found between the SCI-Grand Total and the single-item evaluation form given in that class ( $r = .86$ ). This outcome appears to indicate that the type of instrument used for evaluation makes little difference in the rating of the instructor.

Moderately high correlations were found for each of the subsets of the Grand Total of Student-Constructed Items and the single-item evaluation form given in that class (Table 4.4). The lowest agreements occurred between the SCI Common Item evaluation and the KSU evaluation ( $r = .47$ ), and the SCI Common Item evaluation and the single-item evaluation given with the KSU evaluation ( $r = .41$ ). The low agreement between these measures of evaluation probably occurred in part because the Kansas State University evaluation assesses a large variety of characteristics of

the instructor, whereas the Common Items describe specific teaching skills.

A moderately high correlation occurred between the SCI-Grand Total and the KSU evaluation ( $r = .69$ ) which probably reflected the more global nature of these two measures. This result appears to indicate substantial consistency of evaluation between different classes of the same instructor even considering the above evidence of interclass differences.

In order to test the amount of agreement between the two classes of the instructors participating in the study, the single item evaluation form, which was common to both classes, was used as a basis for analysis.  $t$  tests were used to determine whether significant differences existed between each of the means for this evaluation measure for each of the sixty instructors. Significant differences were found for only twelve instructors ( $p < .05$ ). The correlation matrix for the eight measures of evaluation was re-calculated deleting the twelve instructors. This adjustment resulted in higher correlations for all evaluation measures. As expected the correlation between the two single item evaluations increased (.62 to .84). The correlation between the common items and the KSU-10 increased from .41 to .55 (Table 4.5). In view of the fact that deletion of the 12 cases did not change the basic relationship in the matrix, separate analyses deleting the twelve instructors were not performed.

Table 4.5

Correlation Matrix for Eight Measures of Evaluation  
With Deletion of Twelve Instructors Who Had  
Significantly Different Means on The Two  
Single Item Evaluation Forms

Evaluation Measures	1	2	3	4	5	6	7	8
(1) KSU	--							
(2) KSU-10	.8191	--						
(3) SCI-10	.6781	.8424	--					
(4) Common	.5559	.5499	.6395	--				
(5) Common/Specific	.7130	.7258	.7626	.8457	--			
(6) Specific	.6721	.6640	.6402	.6848	.9029	--		
(7) Grand Total	.7596	.8643	.8732	.8086	.8924	.7620	--	
(8) Idiosyncratic	.5826	.6644	.7223	.6679	.6410	.5127	.8527	--

The Kendall coefficient of concordance ( $W$ ) was used to further test the extent of association among the rank ordering of the sixty instructors on the eight measures of evaluation. Using the two single item measures, six combinations of evaluation measures were tested for the degree of association. The six combinations were chosen to illustrate a variety of combinations of overlapping and non-overlapping measures (Figure 4.2).

The first two tests used six of the eight evaluation measures. Both of the single-item evaluation measures were used in these tests. The Common/Specific Item rankings and the SCI-Grand Total Item rankings were omitted from Test One, and the ranks from the Common Items and the Specific Items were omitted from Test Two. The results of these two tests yielded highly significant values for  $W$ , .73 and .67 respectively, indicating a substantial degree of agreement between these combinations of evaluation measures (Table 4.6).

The second combination of evaluation measures used five measures. Rankings from the single-item evaluations form given in the class with the Kansas State Evaluation Form were omitted. The Specific and Idiosyncratic rankings were also omitted from Test Three. For Test Four, the Common/Specific and the Grand Total rankings plus the Kansas State University-10 were omitted. This test also yielded highly significant values for  $W$ , .77 and .68 respectively (Table 4.7).

Test No.	KSU	KSU-10	SCI-10	Common	Specific	Common/ Specific	Idio- syncratic	SCI Grand Total
1	x	x	x	x	x		x	
2	x	x	x	x		x		x
3	x		x	x		x		x
4	x		x	x	x		x	
5	x	x		x	x		x	
6	x	x		x		x		x

Figure 4.2

Combinations of Evaluation Measures For The  
Kendall Coefficient of Concordance

Table 4.6

Kendall Coefficient of Concordance For  
Six Measures of Evaluation

Test No.	Combinations of Evaluation Measures	df	$\chi^2$	W
(1)	KSU, KSU-10, SCI-10, Common Items, Specific Items, Idiosyncratic Items	59	258.8391	.73118
(2)	KSU, KSU-10, SCI-10, Common Items, Common/Specific Items, SCI-Grand Total	59	235.5475	.66539

\*p < .01

Table 4.7

Kendall Coefficient of Concordance For  
Five Measures of Evaluation

Test No.	Combinations of Evaluation Measures	df	$\chi^2$	W
(3)	KSU, SCI-10, Common Items, Common/Specific Items, SCI-Grand Total	59	226.2758	.76704
(4)	KSU, SCI-10, Common Items, Specific Items, Idiosyncratic Items	59	200.4757	.67958
(5)	KSU, KSU-10, Common Items, Specific Items, Idiosyncratic Items	59	195.550	.66290
(6)	KSU, KSU-10, Common Items, Common/Specific Items, SCI-Grand Total	59	218.0241	.73906

\*p < .01

The third combination of tests also used five measures of evaluation. For this combination, the single-item evaluation forms given with the Student-Constructed Items was omitted from the comparisons. For Test Five, the Common/Specific and SCI-Grand Total measures were also omitted. In Test Six, the Specific Items and the Idiosyncratic Items were omitted in addition to the SCI-10. Results from these two tests were similar to results discussed previously. The values for  $W$  were .66 and .74, respectively, which indicated a high degree of agreement between the measures of evaluation for the sixty instructors. It may be concluded from this analysis that students tend to evaluate the instructor consistently on all measures of evaluation.

A final step in data analysis applied hierarchical groupings using Johnson MAX procedure (Baker, 1972:345-356) to determine whether patterns of ranks could be associated with personal or professional characteristics of the instructors. Hierarchical grouping algorithms can be described as iterative procedures which group those subjects whose measured characteristics are most alike. Rankings from the eight evaluation measures were used to hierarchically group the instructors. At the beginning each instructor was considered to be a group or a weak cluster. (The process proceeds from a weak clustering to a strong clustering.) At each iteration a single new group was formed. The iterative procedure was continued until all instructors formed a single group or strong cluster. The number of iterations

necessary to proceed from the weak to the strong cluster was one less than the number of instructors ( $60 - 1 = 59$ ). The result of each grouping procedure produced a diagram known as a dendogram. The dendogram provides a visual presentation of group formation and the pattern of mergers which lead to the final group. (Figure 4.3).

Hierarchical groups were used to identify rank profiles of instructors. As illustrated in Figure 4.4, five rank profiles were identified. Average ranks for each group on each measure of evaluation are shown in Table 4.8. The first profile (formed at iteration 47) was labeled A-1, Analytical Subject Areas, which included seven instructors: 1, 31, 15, 28, 5, 10, and 30. Four instructors from Radford College and three from Virginia Polytechnic Institute and State University made up this group. Only one of the seven instructors in this group was female. The probability of this event occurring by chance (given the proportion of females in the sample) was .075. There were no instructors from the community college in this group. The probability of this occurring by chance was .058. The group could best be described as males with doctorates, teaching senior college courses that were analytical in nature. Courses taught by these instructors were biology, philosophy, and mathematics. The average rankings received by instructors in this group were generally low. However, they were slightly higher than those in Group Two (Figure 4.4).

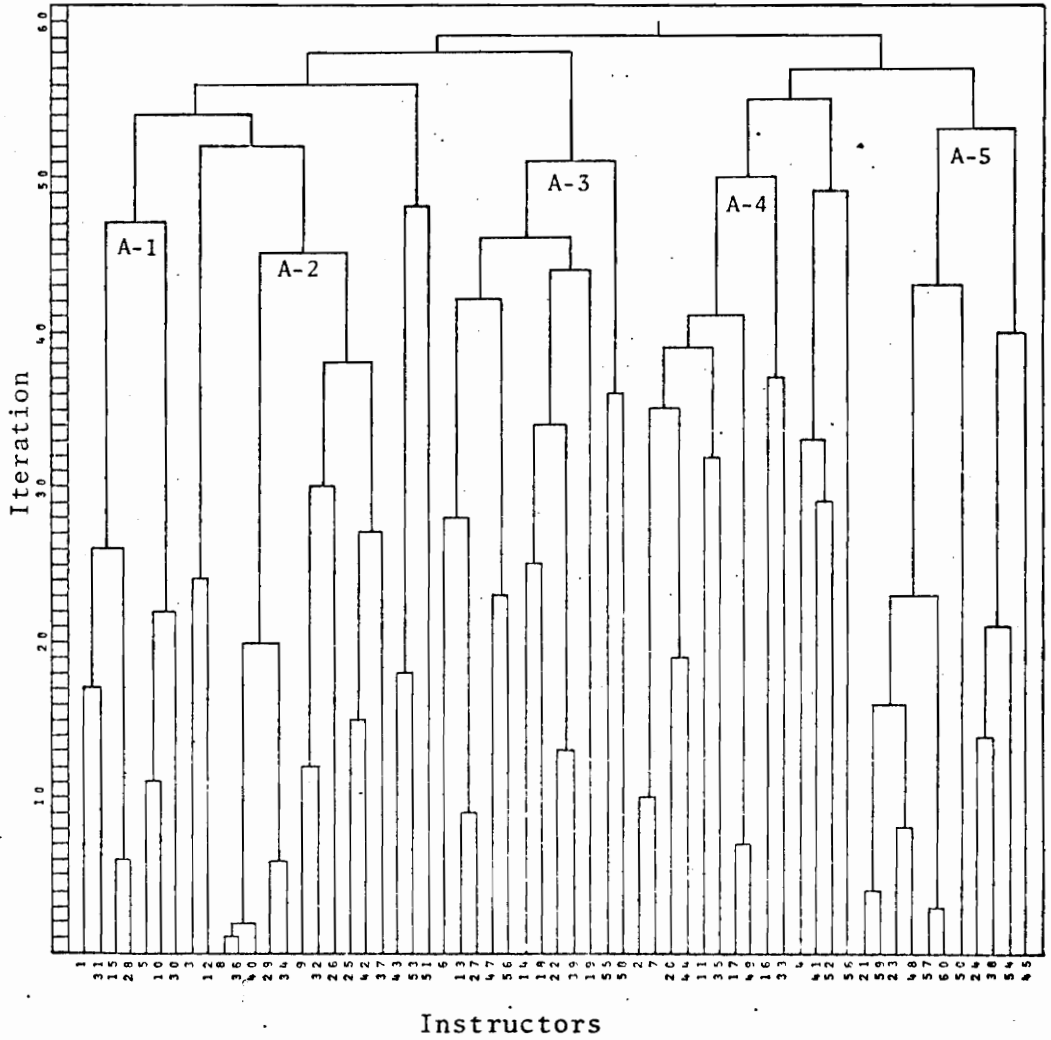


Figure 4.3

Dendrogram for Sixty Instructors Grouped  
by the MAX Procedure

Table 4.8

Average Ranks of Groups For The  
Eight Measures of Evaluation

Group No.	N	Evaluation Measure							
		KSU	KSU-10	SCI-10	Comm.	Comm/ Specific	Specific	G.T.	Idio- Syncratic
A-1	7	14.0	13.1	13.7	36.0	26.4	24.6	19.0	22.4
A-2	11	10.4	8.8	10.0	9.6	7.4	9.6	7.1	11.3
A-3	12	38.4	37.7	25.8	21.9	26.2	27.1	28.5	21.9
A-4	10	32.0	25.6	41.2	45.5	44.0	40.7	45.7	40.8
A-5	11	48.9	51.4	51.4	46.9	50.1	49.0	53.5	54.1

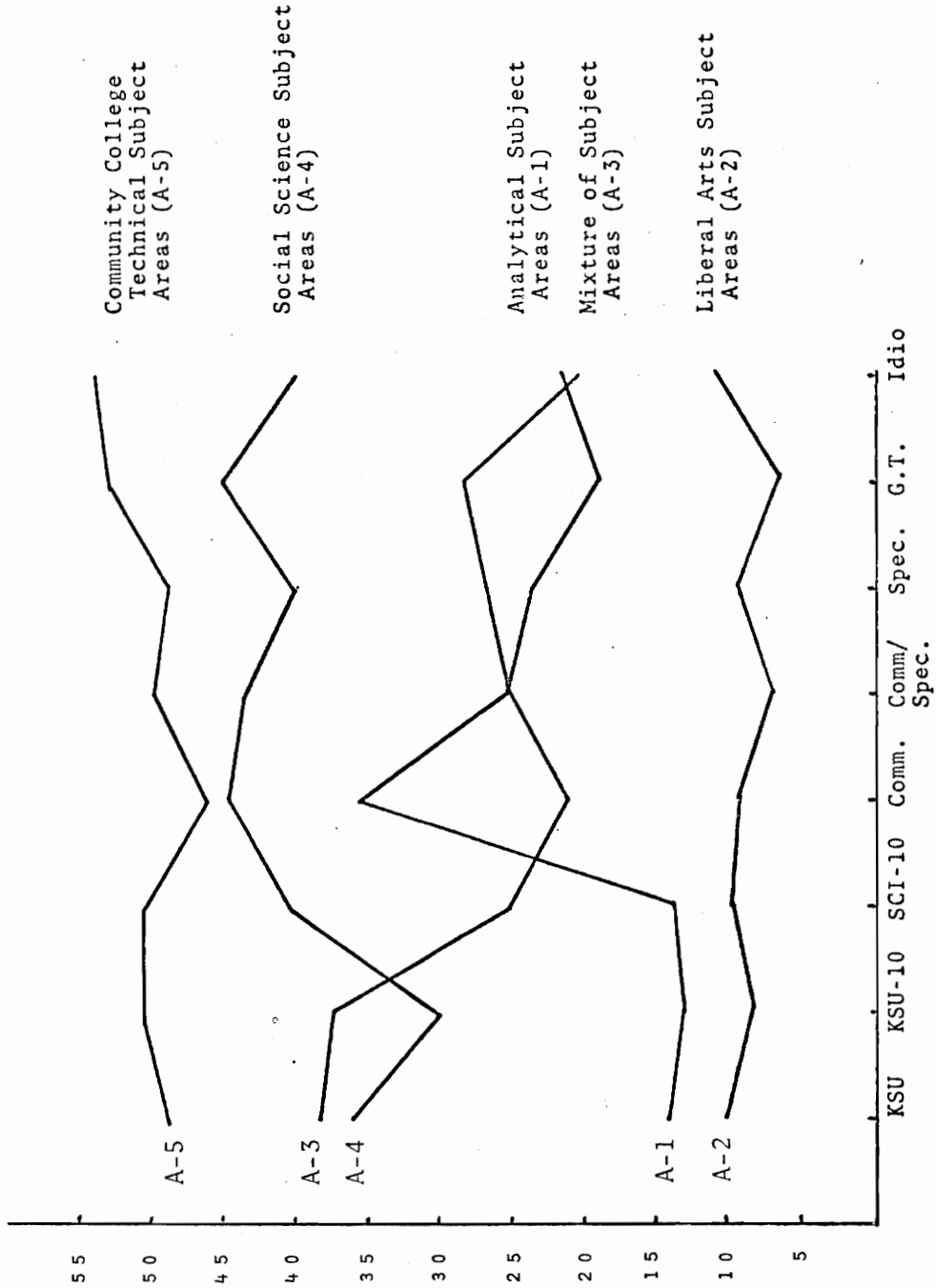


Figure 4.4

Rankings of Instructors Identified From Pattern Analysis for the Eight Measures of Evaluation

The second profile (formed at iteration 45) was labeled A-2, Senior College Liberal Arts Subject Areas. This profile included eleven instructors: 8, 36, 40, 29, 34, 9, 32, 26, 25, 42, and 37. There were six female and five male instructors in this group. Only one community college instructor appeared in this group. The probability of this event occurring by chance was .021. Instructors appearing in this group were predominately those teaching liberal arts courses in the two senior colleges. Liberal arts teachers have been traditionally described as subject-oriented rather than student-oriented, a fact which could account for the low rankings received by this group. Experience and degree characteristics could be important for membership in these groups. Instructors for groups A-1 and A-2 had an average of less than six years' teaching experience and held the doctorate in almost every case. Instructors 3 and 12, who would have entered A-2 at iteration 52, had profile patterns which did not closely resemble those of others in A-2.

Three instructors--43, 53, and 51--formed a group (at iteration 48) which could not be identified with either Profile A-1 or Profile A-2.

A middle group, the third profile (formed at iteration 51), was labeled A-3, Mixture of Subject Areas. This group, composed of three female and nine male instructors, was defined as one which did not show predominate characteristics in any of the personal profile areas.

Mean rankings for evaluations were close to the median for the five profiles.

The fourth profile (formed at iteration 50) was label A-4, Social Science Subject Areas, and included ten instructors: 2, 7, 20, 44, 11, 35, 17, 49, 16, and 33. This group included three female and seven male instructors. Subject areas taught by these instructors were mostly from the social science and fine arts areas. The majority of instructors in this profile held Master's degrees and had over six years' teaching experience. The social studies and fine arts areas are usually considered to be even more student-oriented than the natural sciences.

Four instructors--4, 41, 52, and 46--formed a group (at iteration 49) which could not be identified with Profile A-4.

A fifth profile labeled A-5, Community College Technical Subject Areas (formed at iteration 53), included eleven instructors: 21, 59, 23, 48, 57, 60, 50, 24, 38, 54, and 45. Community college instructors made up seven of the eleven members. The majority of the instructors held master's degrees with over six years' teaching experience. It should be noted that these characteristics are important elements of the community college emphasis on student-oriented teaching.

The five groups identified from this analysis were inspected for the placement of the twelve instructors who were found to have significant differences in the means on

the two single-item evaluation forms. These instructors were spread throughout the groups and non-groups. It was therefore determined that these twelve instructors did not substantially alter the groupings. Had a majority of the twelve instructors appeared in a single group, it would have been necessary to perform the analysis again, deleting these instructors.

## Chapter 5

### CONCLUSIONS AND RECOMMENDATIONS

This study examined differences between college instructors based on student input using three types of evaluative instruments. Instructors have frequently expressed preferences for one instrument over another as a means of obtaining student evaluative input. Regardless of the instrument used, administrators usually desire a single score to summarize the evaluation of each instructor.

This study posed four questions specific to these areas of administrator and instructor concern with respect to evaluation. Each question is stated below along with the conclusions reached.

1. To what extent is it justifiable to distinguish among instructors on the basis of summated ratings?

For the vast majority of faculty members, practical differences cannot be determined on the basis of summated scores. Results from eight measures of evaluation analyzed using the Duncan New Multiple Range Test show that extensive overlap of equivalent subgroups makes it impossible to distinguish between one instructor and another unless the ranks of their means are substantially different.

The writer concludes that differences between instructors cannot be determined on the basis of summated

scores except in cases of extremes where an instructor would receive a very high or a very low rating. It is suggested that an alternative method such as a ranking system described in chapter 4, p. 44 be used if requests are made for differentiating between instructors.

2. What questionnaire item topics are indicated by students as basic to the ratings of all or nearly all instructors?

Five item topics were identified from the SCI evaluation form as basic to the ratings of all or nearly all instructors.

1. Knowledge of material.
2. Interesting presentation.
3. Teaching method--ability to explain material.
4. Grading fairness.
5. Oriented toward or relates to students.

These items may be described as those factors which students relate directly to the learning process and to the way they view the instructor as oriented to the student.

The researcher interprets these items as defining three areas of major student concern for instruction.

1. That the faculty member knows the subject.
2. That he presents the subject effectively.
3. That the instructor relates to the student in a personal and a professional manner.

Limitation: A question may be raised concerning these five common items as they relate to the groups identified from the pattern analysis. For example, does grading fairness have the same meaning for group A-1 as it does for group A-3? A disproportionate representation of these items on the summary sheet would imply non-comparability on the measurements from one instructor to another. Inspection of the summary sheets for each instructor indicated that the five common items were not disproportionately distributed. Because of the method of defining specific and idiosyncratic items, the proportion of these items varied from instructor to instructor. Common items appeared to have proportionate consistency across all instructors.

3. To what extent do summated ratings based on student-constructed questionnaire items correlate with those obtained from other evaluation instruments?

High values for Kendall's Coefficient of Concordance were obtained for analyses involving logically selected subsets of all eight evaluation measures, indicating a high degree of association between the different measures of evaluation. From the correlations obtained from the eight evaluation measures, the highest positive correlation (.86) was between the total for the student-constructed items and the single-item form administered in the same classes. Moderately high correlations were found between the specific items and the idiosyncratic items with the two objective types of evaluation. The common items correlated less well

with the other types of evaluation than the previously mentioned subsets of items. The common items stressed factors that are basic to the teaching process while all areas of evaluation are considered in the single item evaluation and the Kansas State University evaluation. The contrast is that the specific and idiosyncratic items were similarly diverse.

From the relatively high correlations between rankings on the eight measures of evaluation, the researcher concludes that the instructor will receive basically the same ranking no matter which instrument is used for evaluation. As was noted in the first question, it is the scoring method that is most significant in the evaluation process rather than the type of instrument that is used.

4. Do instructor profiles based on summated ratings from student-constructed items and other rating scales provide a basis for distinguishing among instructors on the basis of subject matter, specialization, degree level, experience, sex, and age?

The use of pattern analysis identified five groups which had similar characteristics based on rankings obtained from the different evaluation measures. Those faculty members receiving the highest evaluations were found in the areas of the social sciences and vocational courses. In contrast, instructors of courses in the liberal arts area, which are traditionally considered to be highly subject-oriented, received the lowest evaluations.

It may be concluded from these findings that the instructor's teaching specialty does make a difference in the kind of evaluation he will receive. Instructors in courses whose immediate application is evident will receive higher ratings than those who teach courses whose content has long range application.

Instructors in the analytical subject areas scored low on the standardized form and the two single item evaluation forms but contrastingly high on the student constructed common items. Two conclusions may be drawn from this contrast in ratings. First, students from these classes do not hesitate to rate instructors low on an objective form, but if forced to identify items specifically related to the teaching/learning process will evaluate the instructor differently. A second conclusion that may be drawn from this contrast is that the common items do not identify the source of discontent the student has with the instructor. The concerns of the student appear to relate to more general characteristics of the instructor. Thus it appears from this analysis that the subject areas taught do make a difference in evaluation results.

Another contrast was evident with respect to instructors holding the doctorate who had little or no teaching experience. These teachers tended to receive lower ratings than those instructors who held only the Master's degree but had six to ten years' experience in higher education.

No grouping comparisons could be made for instructors with the doctorate and over five years of teaching experience or for those instructors who held the master's degree and had little or no experience.

### Recommendations

As a result of this study and the conclusions drawn, the following recommendations for further study are made.

Since ratings of instructors seemed to vary according to the courses taught, a study should be made to attempt to determine whether differences between ratings stem from a pre-conceived bias for the course on the part of the student or whether teaching methods peculiar to particular subjects cause some instructors to be rated low while others always receive high ratings. Determining the preference of students for certain courses and subjects might be one method of conducting this study. If predictions of ratings for certain courses can be made, then formulas can be developed which will adjust an instructor's rating for the courses he is required to teach.

For the present study, groups were defined by certain personal and professional characteristics. A study which examines other personal and professional characteristics of instructors might be useful in further defining faculty groupings.

Perhaps the most important recommendation for further study is one related to the instructors themselves.

When asked to react to evaluation, instructors tend to react to the type of instrument used. Findings from this study show a consistency of ratings by students no matter what instrument is used. The important factor is the relative position of the instructor to other instructors after the analysis of scores has been completed. Findings from this study indicated that no discrimination could be made between instructors whose scores tended to cluster in the middle group.

As a result of the conclusions drawn from this study, further analysis of the relationship of experience and degree held to ratings received by instructors on student evaluation is also recommended.

## REFERENCES

## REFERENCES

- Baker, Frank B. "Numerical Taxonomy for Educational Researchers," Review of Educational Research, 42 (1972), 345-356.
- Beck, Nancy. Telephone interview. November 20, 1975.
- Boyce, Arthur C., ed. "Methods in Measuring Teachers' Efficiency," Fourteenth Yearbook, National Society for the Study of Education. Chicago: Public School Publishing Company, 1920.
- Bryant, P. T. "By Their Fruits Ye Shall Know Them," The Journal of Higher Education, 38 (1967), 326-330.
- Centra, John A. "College Teaching: Who Should Evaluate It?" Findings, Educational Testing Service (Winter 1974), pp. 5-8.
- Cohen, Arthur M., and Florence B. Brawer. "Evaluating Faculty," Change, 4 (1972), 33a-33c.
- Conover, W. G. Practical Nonparametric Statistics. New York: John Wiley & Sons, Inc., 1971.
- Costin, Frank, W. T. Greenough, and Robert J. Menges. "Student Ratings of College Teaching: Reliability, Validity, and Usefulness," Review of Educational Research, 42 (1971), 511-531.
- Dixon, W. J., ed. BMD-Biomedical Computer Programs. 3d ed. Los Angeles: University of California Press, 1973.
- Doyle, Kenneth O., Jr., and Susan E. Whitely. "Student Ratings as Criteria for Effective Teaching," American Educational Research Journal, 11 (1974), 259-274.
- Eble, Kenneth E. Professors as Teachers. San Francisco: Jossey-Bass, Inc., 1972.
- Eble, Kenneth E. "What Are We Afraid Of?" College English, January 1974, pp. 451-456.

- Edwards, Allen R. Experimental Design in Psychological Research. 4th ed. New York: Holt, Rinehart and Winston, Inc., 1972.
- Epstein, Charlotte. "Students Grade Teachers?" Community and Junior College Journal, 44 (1974), 32-33.
- Frey, Peter W. "Comparative Judgment Scaling of Student Course Ratings," American Educational Research Journal, 10 (1973), 149-154.
- Gage, N. L., ed. Handbook of Research on Teaching. Chicago: Rand McNally, 1963.
- Gessner, Peter K. "Evaluation of Instruction," Science, 180 (1973), 566-569.
- Grasha, Anthony F. Evaluating Teaching: Some Problems. Institute for Research and Training in Higher Education, Cincinnati University, Ohio. 4, No. 3 (Fall 1972), pp. 1-5 (Microfilm).
- Greenwood, Gordon E., and others. "Student Evaluation of College Teaching Behaviors Instrument: A Factor Analysis," Journal of Higher Education, 44 (1973), 597-604.
- Hoyt, Donald P. "Identifying Effective Instructional Procedures," Improving College and University Teaching. (Winter, 1973a), pp. 73-76.
- Hoyt, Donald P. "Measurement of Instructional Effectiveness," Research in Higher Education, 1 (1973b), 367-378.
- Hoyt, Donald P. Using the Report to Faculty Members to Improve Instruction," Manual for Interpretation of Printouts for Kansas State University Evaluations Processed by the University, February, 1970c.
- Kent, Laura. "Student Evaluation of Teaching," Educational Record, Summer 1966, pp. 376-403.
- Kiernan, Irene R. "Student Evaluation Re-evaluated," Community and Junior College Journal, 45 (1975), 25-27.
- Klugh, Henry E. Statistics: The Essentials For Research. New York: John Wiley & Sons, Inc., 1970.
- Kramer, C. Y. "Extension of Multiple Range Tests to Group Means With Unequal Numbers of Replications," Biometrics, 12 (1956), 302-310.

- Lee, Calvin B. T., ed. Improving College Teaching. Washington, D. C.: American Council on Education, 1967.
- Maiher, Anne D. "'Opening Up' Faculty Evaluation," Regional Spotlight, Southern Regional Education Board, September 1975, pp. 1-6.
- Matthei, Richard K., and Henry R. Winstock. "Evaluation of Instruction--An Elusive Target," School Research Information Service, Phi Delta Kappa, (Winter 1973), pp. 20-21.
- Miller, Richard I. Evaluating Faculty Performance. San Francisco: Jossey-Bass, Inc., 1972.
- Musella, Donald, and Reuben Rusch. "Student Opinion on College Teaching," Improving College and University Teaching, 16 (1968), 137-140.
- Nelson, C. Van, and Elaine R. Cossart. "The Adaptation of the Method of Summated Ratings Using Normal Deviate Weights to a Form for Student Evaluation of Instructors" (paper presented at the meeting of the American Educational Research Association, Washington, D. C., April, 1974), pp. 1-13. (Mimeographed).
- Pogue, F. G., Jr. "Student Ratings of the Ideal Teacher," Improving College and University Teaching, 15 (1967), 133-136.
- Rodin, Miriam, and Burton Rodin. "Student Evaluations of Teachers," Science, 177 (1972), 1164-1166.
- Rowland, Ray. "Can Teaching Be Measured Objectively?" Improving College and University Teaching, 18 (1970), 153-157.
- Siebert, Warren F. Telephone Interview. November 20, 1975.
- Siebring, B. R., and M. E. Schaff. "A Survey of the Literature Concerning Student Evaluation of Teaching" (paper presented at the Symposium on Student Evaluation of Teaching sponsored by the Division of Chemical Education at the 165th National meeting of ACS, Dallas, Texas, April, 1973).
- Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.

- Subboviah, Michael J., and Joel R. Levin. "Determining the Characteristics of the Ideal Professor: An Alternative Approach" (paper presented at the meeting of the American Educational Research Association, Washington, D. C., April, 1974).
- Walker, Billy D. "Course Difficulty and Student Rating of Teaching," Improving College and University Teaching, 22 (1974), 18-20.
- Weaver, Carl H. "Instructor Rating by College Students," Journal of Educational Psychology, 51 (1960), 21-25.
- Wennstrom, Jan T., and T. W. Rigby. An Analysis of Faculty Evaluation Procedures in Accredited Community and/or Junior Colleges in the United States. (Abstract), Baylor University, 1974, pp. iii - xi.
- Whitely, Susan E., Kenneth O. Doyle, Jr., and Karen Hopkinson. Student Ratings and Criteria for Effective Teaching, Measurement Services Center, Report 731F, University of Minnesota, October, 1973, pp. 1-41.

APPENDICES

Appendix A

Questionnaire To Instructors Participating  
in Evaluation

Instructor Number \_\_\_\_\_

Classes \_\_\_\_\_

College \_\_\_\_\_

QUESTIONNAIRE TO INSTRUCTORS PARTICIPATING  
IN EVALUATION

For descriptive purposes of the evaluation study it is necessary that we describe faculty participating in the study.

1. Indicate the number of years you have been employed by this college.
  - \_\_\_\_\_ 1. 1 - 3 years
  - \_\_\_\_\_ 2. 4 - 5 years
  - \_\_\_\_\_ 3. 6 - 10 years
  - \_\_\_\_\_ 4. Over 10 years
  
2. Indicate your total number of years of teaching experience. (Include public school and college experience.)
  - \_\_\_\_\_ 1. Less than 3
  - \_\_\_\_\_ 2. 4 - 5 years
  - \_\_\_\_\_ 3. 6 - 10 years
  - \_\_\_\_\_ 4. Over 10 years
  
3. What is your highest earned degree?
  - \_\_\_\_\_ 1. Diploma
  - \_\_\_\_\_ 2. Bachelors
  - \_\_\_\_\_ 3. Masters
  - \_\_\_\_\_ 4. Doctorate
  
4. In what field is your highest degree?
 

---
  
5. Indicate your age classification.
  - \_\_\_\_\_ 1. 20 - 30
  - \_\_\_\_\_ 2. 31 - 40
  - \_\_\_\_\_ 3. 41 - 50
  - \_\_\_\_\_ 4. Over 50
  
6. Sex: \_\_\_\_\_ 1. Female \_\_\_\_\_ 2. Male
  
7. Do you wish to receive results of the evaluation by your classes?
  - \_\_\_\_\_ 1. Yes
  - \_\_\_\_\_ 2. No

Appendix B

Procedures For Instructors Participating  
in Student Evaluation Project

## PROCEDURES

T0: Instructors Participating in Student Evaluation Project

Attached are two evaluation sets, a cassette tape, the text of the cassette tape, and a short questionnaire for you. Please keep all evaluation materials together and separated by classes.

You should return all materials to \_\_\_\_\_  
by \_\_\_\_\_.

Please remember that all evaluations are confidential and that identification is for return purposes only. Results will be used only for statistical analysis. If you wish, you may receive analysis of the results of your classes only.

Instructions for administering evaluations in your classes are found on the front of each packet.

Thank you for your assistance in this project. If you have questions, please call me at 674-4121 or 639-1263.

Edith H. Carter

CLASS \_\_\_\_\_

INSTRUCTIONS FOR ADMINISTERING EVALUATION  
Standardized Evaluation

Directions:

1. Each student packet contains all needed evaluation materials.  
Materials should not be torn apart.
2. Pass out packets to all students. Have students read the instructions through Step I.
3. Ask students to turn to the first evaluation form in their package and mark with pencil the first 58 items.
4. When everyone has finished, have students read Step II of the directions.
5. Ask class to turn to Teacher Evaluation Questionnaire. Remind students to complete only question 3 on this form.
6. Do not separate evaluation forms.
7. Have a student put all evaluation materials in attached envelope and seal the envelope.
8. You have now completed the evaluation process for this class.

Thank you.

CLASS \_\_\_\_\_

INSTRUCTIONS FOR ADMINISTERING EVALUATION  
Student Constructed Items

For this evaluation please use the cassette or, if a cassette player is not available, please read the text of the cassette tape at the designated time.

The entire evaluation should take not longer than 30 minutes.

Directions:

1. Each student packet contains all needed evaluation materials. Materials should not be torn apart.
2. Pass out packets to all students. Have students read the instructions through Step I.
3. Play cassette tape or read to the class the text of the cassette tape.
4. Ask students to turn to the Student Constructed Items Form in their package.
5. Have them write evaluation items. In some cases you may have to encourage students to write items.
6. When everyone has finished, have students read Step II of the directions.
7. Ask class to turn to Teacher Evaluation Questionnaire. Remind students to complete only question 3 on this form.
8. Do not separate evaluation forms.

Instructions for Administering Evaluation  
Page 2

9. Have a student put all evaluation materials in attached envelope and seal the envelope.
10. You have now completed the evaluation process for the class.

Thank you.

## TEXT OF TAPE FOR EVALUATION

Class members, this is a request to each of you to help develop a new type of evaluation system for instructors.

How often have you been given a questionnaire and found that it did not ask the questions that you thought were important?

For example, you go with a friend to a new car dealer and watch her complete a questionnaire on a new car. The friend responds that she likes the color, the design, the vinyl top, the color of the interior and the carpet. When you leave the car dealer she responds that what she really wanted to say about the new car was that she bought it because it had a stereo tape player.

Another example is the hunter who is given a questionnaire when he picks up his hunting license. The questions go something like this: Do you like to hunt? He responds yes. Do you hunt often? He responds yes. Do you like to hunt deer? He responds no. Do you like to hunt small animals? He responds no. After turning in the questionnaire he remarks that he wishes they had asked if he liked to hunt birds and to go on weekend hunting trips.

Many of us have the same problem when we do faculty evaluations. We have answers for which there are no questions. For the evaluation that you are about to do, we want you to write the questions for the instructor and the instructional methods you have observed in this class.

For example, something that would be important to you about an instructor may be an item such as this: "The instructor stands while teaching."

Text of Tape for Evaluation  
Page 2

After you have written an item, evaluate the instructor on that item by marking always, usually, sometimes, or never in the appropriate column on your sheet.

Another example may be: "The instructor gives good examples." Again, you would mark the item always, usually, sometimes, or never.

You will then continue to write and evaluate the instructor on as many items as you feel are important to describe the instructor.

You should now proceed with writing your evaluation items. Areas that you may want to consider for evaluation are: personal and professional characteristics, student-teacher relationships, communication, classroom management, and student achievement.

Appendix C

Instructions To Students For  
Student Constructed Items

INSTRUCTIONS FOR EVALUATION PROCESS  
Student Constructed Items

To All Students:

You can help us to help you by participating in a study involving faculty evaluation. The results of this evaluation will not be used as a basis for tenure or salary for your instructor but may be used for improvement of instruction. You are requested to complete the two steps of the evaluation as accurately as possible. Your answers will be confidential and will be used for data analysis only. The entire evaluation should take no longer than 30 minutes.

STEP I

On the first sheet in your packet please write as many evaluation items as you feel are necessary to evaluate the instruction you have received in this class. Items should be written so that if the items were true it would compliment the instructor. When you have written as many items as you feel are necessary please mark each item with one of the following ratings:

Always, Usually, Sometimes, Never.

For example, an item may be: "The instructor arrives at class on time." Since the instructor frequently arrives at class late, you would mark this item "Sometimes."

YOUR INSTRUCTOR HAS A TAPE OR WILL READ YOU OTHER EXAMPLES OF ITEMS

(Continued on Page 2)

Instructions for Evaluation Process  
Student Constructed Items  
Page 2

STEP II

Please refer to the second sheet in your packet which is labeled Teacher Evaluation Questionnaire. On this form you are asked to rate the instructor on a 1 - 10 scale with 1 being the lowest rating and 10 being the highest rating. Your evaluation should be based on a comparison with other teachers you have had at this college. Please see form for other evaluation criteria. (Complete question 3 only)

You have now completed the evaluation set. Thank you for your participation.

## STUDENT CONSTRUCTED ITEMS FORM

## Instructions:

Items should be written so that if the item were true it would compliment the instructor. Write as many items as necessary to evaluate the instruction you have received in this class. Then circle always, usually, sometimes, or never for each item.

Always Usually Sometimes Never 1. \_\_\_\_\_

Always Usually Sometimes Never 2. \_\_\_\_\_

Always Usually Sometimes Never 3. \_\_\_\_\_

Always Usually Sometimes Never 4. \_\_\_\_\_

Always Usually Sometimes Never 5. \_\_\_\_\_

Always Usually Sometimes Never 6. \_\_\_\_\_

Always Usually Sometimes Never 7. \_\_\_\_\_

Always Usually Sometimes Never 8. \_\_\_\_\_

Always Usually Sometimes Never 9. \_\_\_\_\_

Always Usually Sometimes Never 10. \_\_\_\_\_

Always Usually Sometimes Never 11. \_\_\_\_\_

Always Usually Sometimes Never 12. \_\_\_\_\_

TEACHER EVALUATION QUESTIONNAIRE

DO NOT WRITE IN THIS SPACE

PROFESSOR \_\_\_\_\_ QUARTER \_\_\_\_\_

DEPT \_\_\_\_\_ COURSE \_\_\_\_\_ YEAR \_\_\_\_\_

- MAKE ALL MARKS HEAVY AND BLACK
- MAKE ALL ERASURES CLEAN
- MAKE NO STRAY MARKS
- USE ONLY A NO. 2 PENCIL
- FILL SPACES BETWEEN DOTTED LINES COMPLETELY

1. YOUR OVERALL GRADE POINT AVERAGE

4.00	3.00	2.00	1.00	0.00
3.50	2.50	1.50	0.50	
3.00	2.00	1.00	0.00	
2.50	1.50	0.50		
2.00	1.00	0.00		
1.50	0.50			
1.00	0.00			
0.50				
0.00				

2. GRADE WHICH YOU EXPECT TO RECEIVE IN THIS COURSE.

A	B	C	D	E	F	OTHER
---	---	---	---	---	---	-------

3. EVALUATE YOUR TEACHER IN THIS COURSE IN COMPARISON WITH OTHER TEACHERS YOU HAVE HAD AT NRCC BY MARKING ONE NUMBER BETWEEN 1 AND 10 BELOW. CRITERIA TO BE CONSIDERED IN YOUR EVALUATION SHOULD INCLUDE (BUT NOT NECESSARILY BE LIMITED TO) THE FOLLOWING: EFFECTIVENESS IN PRESENTATION OF SUBJECT MATTER; APPARENT KNOWLEDGE OF SUBJECT MATTER; FAIRNESS IN GRADING AND DEMANDS FOR OUTSIDE WORK; COMMITMENT AND ENTHUSIASM; AVAILABILITY TO STUDENTS.

10	9	8	7	6	5	4	3	2	1
EXCELLENT	GOOD		FAIR			POOR			

4. BRIEFLY DESCRIBE BELOW SOME OF REASONS FOR YOUR EVALUATION AS INDICATED ABOVE.

TEACHER IDENTIFICATION NUMBER

<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9

DEPARTMENT

<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9

COURSE INDEX NUMBER

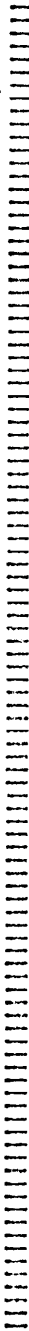
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9

QUARTER

<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9

YEAR

<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9
<input type="checkbox"/>	0	1	2	3	4	5	6	7	8	9



Appendix D

Instructions To Students For  
Standardized Evaluation

INSTRUCTIONS FOR EVALUATION PROCESS  
Standardized Evaluation

To All Students:

You can help us to help you by participating in a study involving faculty evaluation. The results of this evaluation will not be used as a basis for tenure or salary for your instructor but may be used for improvement of instruction. You are requested to complete the two steps of the evaluation as accurately as possible. Your answers will be confidential and will be used for data analysis only. The entire evaluation should take no longer than 30 minutes.

STEP I

Please complete the first 58 items found on Part I of the first evaluation form. You do not need to complete the other parts of the evaluation form. Please use pencil to mark your answers.

STEP II

Please refer to the second sheet in your packet which is labeled Teacher Evaluation Questionnaire. On this form you are asked to rate the instructor on a 1 - 10 scale with 1 being the lowest rating and 10 being the highest rating. Your evaluation should be based on a comparison with other teachers you have had at this college. Please see form for other evaluation criteria. (Complete question 3 only)

You have now completed the evaluation set. Thank you for your participation.



**TEACHER EVALUATION  
QUESTIONNAIRE**

DO NOT WRITE IN THIS SPACE

PROFESSOR \_\_\_\_\_ QUARTER \_\_\_\_\_

DEPT \_\_\_\_\_ COURSE \_\_\_\_\_ YEAR \_\_\_\_\_

- MAKE ALL MARKS HEAVY AND BLACK
- MAKE ALL ERASURES CLEAN
- MAKE NO STRAY MARKS
- USE ONLY A NO. 2 PENCIL
- FILL SPACES BETWEEN DOTTED LINES COMPLETELY

**1.  
YOUR OVERALL GRADE POINT AVERAGE**

0--	1.00-	2.00-	3.00-	4.00-
1.99	2.49	2.99	3.49	4.00

**2.  
GRADE WHICH YOU EXPECT TO RECEIVE IN THIS COURSE.**

A	B	C	D	E	F	OTHER
---	---	---	---	---	---	-------

**3.**  
EVALUATE YOUR TEACHER IN THIS COURSE IN COMPARISON WITH OTHER TEACHERS YOU HAVE HAD AT NRCC BY MARKING ONE NUMBER BETWEEN 1 AND 10 BELOW. CRITERIA TO BE CONSIDERED IN YOUR EVALUATION SHOULD INCLUDE (BUT NOT NECESSARILY BE LIMITED TO) THE FOLLOWING: EFFECTIVENESS IN PRESENTATION OF SUBJECT MATTER; APPARENT KNOWLEDGE OF SUBJECT MATTER; FAIRNESS IN GRADING AND DEMANDS FOR OUTSIDE WORK; COMMITMENT AND ENTHUSIASM; AVAILABILITY TO STUDENTS.

[10]	[9]	[8]	[7]	[6]	[5]	[4]	[3]	[2]	[1]
EXCELLENT		GOOD			FAIR		POOR		

**4.**  
BRIEFLY DESCRIBE BELOW SOME OF REASONS FOR YOUR EVALUATION AS INDICATED ABOVE.

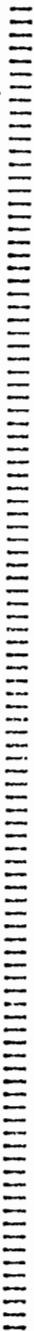
TEACHER IDENTIFICATION NUMBER

DEPARTMENT

COURSE INDEX NUMBER

QUARTER

YEAR



Appendix E

Instructor Summary Sheet



Appendix F

Examples of Student  
Constructed Items  
and  
Classification of 89  
Student Constructed Items

## EXAMPLES OF STUDENT CONSTRUCTED ITEMS

These items are quoted directly from the evaluation forms submitted by students.

Examples are given for the five classifications of common items and other items. Specific items changed from instructor to instructor.

### Rating Scale

- 4 Always
- 3 Usually
- 2 Sometimes
- 1 Never

#### Rating Knowledge of Material

- 2 Knows subject matter
- 4 Knowledge of course
- 3 Knowledge in subject matter
- 4 Instructor knows what he is talking about
- 4 Has an outstanding knowledge of the material she is teaching

#### Rating Interesting Presentation

- 4 Makes class interesting
- 2 Makes class enjoyable
- 2 Instructor makes the class stimulating and interesting
- 3 Makes lectures interesting

Examples of Student Constructed Items  
Page 2

Rating Teaching Method, Ability to Explain Material

- 3 Presents material in understandable form
- 4 Instructor gives clear precise examples when lecturing
- 1 Explains fully and well
- 4 The teacher thoroughly explains the work assigned to students
- 2 Explains material clearly

Rating Grading Fairness

- 4 Instructor is fair in grading
- 2 Grades fairly and without using bias
- 3 Grades fairly and accurately
- 2 Fair in grading

Rating Oriented Toward or Relates to Students

- 3 Shows concern for his students
- 4 The teacher has a good relationship with students
- 4 The teacher seemed interested in the students
- 4 Interested in you as a person

Rating Interest in Teaching and/or Subject

- 4 Instructor is enthusiastic about his subject
- 3 The instructor teaches because she likes the profession
- 4 Enjoys teaching

## Examples of Student Constructed Items

Page 3

Rating Testing Fairness

- 2 Do the tests cover the material stressed during class periods
- 4 The teacher gives fair and objective tests
- 4 Tests emphasize points emphasized in lecture
- 4 Gave ample notice for oncoming tests
- 3 Gave fair thorough tests

Rating Miscellaneous Items Taken From Other Categories

- 4 The instructor makes clear the objectives of the class
- 2 Uses visual aids effectively
- 4 Instructor has neat appearance
- 4 Excellent personality characteristics
- 3 Instructor maintains proper control of the class
- 1 Instructor will readily accept other reasonable points of view
- 3 Is available when needed
- 4 Class time utilized in best way

CLASSIFICATION OF STUDENT  
CONSTRUCTED ITEMS

1. Knowledge of material
2. Fairness to students
3. Sense of humor
4. Appearance
5. Teaching methods--class discussion
6. Relation of material to other fields
7. Use of audio visual materials in class
8. Adherence to course outline
9. Teaching method - organization of lecture
10. Outside assignments
11. Interesting presentation
12. Organization of material
13. Teaching method--ability to explain material
14. Availability for conferences
15. Variety of teaching methods
16. Stimulation of learning
17. Grading--fairness
18. Interest in teaching and/or subject
19. Clear speech
20. Testing--good test items
21. Testing--fairness
22. Oriented toward or relates to students
23. Provides assistance with required work
24. Efficient use of class time
25. Prompt feedback to students

## Classification of Student Constructed Items

Page 2

26. Clear objectives
27. Presents material slowly enough for comprehension
28. Relates material to application
29. Personality/mood of instructor
30. Testing--flexibility
31. Adds supplementary material
32. Adequately prepares for class
33. Gives good notes
34. Free of annoying mannerisms
35. Unbiased--encourages other points of view
36. Receptive to questions
37. Well-mannered
38. Provides ample break in class period
39. Has control of class
40. Encourages feedback from students
41. Legible handwriting in use of blackboard
42. Comfortable classroom atmosphere
43. Identifies instructional needs
44. Professional attitude
45. Demonstrates and enforces good safety practices
46. Encourages achievement and pride in work
47. Recommend teacher or would take same teacher again
48. Reviews material for tests
49. Provides assistance in labs
50. Creative--intelligent attitude

Classification of Student Constructed Items  
Page 3

51. Liberal attendance policy
52. Requires evaluation
53. Good teacher
54. Other students like class
55. Professor attends class regularly
56. Instructor makes class worthwhile
57. Requires reasonable amount of work
58. Requires the proper level of work
59. Is up to date on subject material
60. Knows school policies
61. Speaks concisely
62. Practices things taught in class
63. Instructor has had work experience
64. Good character
65. Liked or recommended by other students
66. Instructor earns salary
67. Includes personal viewpoints
68. Makes pertinent announcements
69. Modest
70. Admits mistakes
71. Compares favorably with other teachers
72. Honest
73. Has self-confidence
74. Provides for class when absent
75. Knows students' names

Classification of Student Constructed Items  
Page 4

76. Views the subject as important
77. Dependable
78. Makes course worthwhile
79. Respected by students
80. Offers constructive criticism
81. Relates material to application
82. Continues to study in field of specialty
83. Enjoyed teacher and/or course
84. Participates in class activities
85. Handicap inconsequential in instruction
86. Displays a good general knowledge
87. Attempts to see students' viewpoints on assignments
88. Compliments good work
89. Has eye contact with students

Appendix G

College Summary Sheet

College Summary Sheet  
Student Constructed Items

Item Category	Total Responses Per Instructor (Instructor ID Number)				
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
.					
.					
.					
N					
Totals in each cell					
= $\frac{\text{Sum of values}}{\text{No. of items}}$					

Appendix H

Instructor Report For  
Student Constructed Items

Report to Instructors  
Student Constructed Items

Instructor \_\_\_\_\_ College \_\_\_\_\_

- I. Five items were found to be common to all instructors in this study. The common items and the frequency of student responses for each item are reported below.

Common Items	Always (4)	Usually (3)	Sometimes (2)	Never (1)
1. Knowledge of material	—	—	—	—
2. Interesting presentation	—	—	—	—
3. Grading fairness	—	—	—	—
4. Teaching method ability to explain material	—	—	—	—
5. Oriented or related to students	—	—	—	—

Your mean score for common items \_\_\_\_\_

Your rank among the sixty instructors from three different colleges: \_\_\_\_\_

(Ranks are from 1 to 60 with 60 being the highest rank).

- II. Certain items specific to each instructor were identified. The frequency of student responses for each item which is specific for you is reported below.

Specific Items	Always (4)	Usually (3)	Sometimes (2)	Never (1)
_____	—	—	—	—
_____	—	—	—	—
_____	—	—	—	—
_____	—	—	—	—
_____	—	—	—	—
_____	—	—	—	—
_____	—	—	—	—

Your mean score for specific items \_\_\_\_\_

Your rank among sixty instructors \_\_\_\_\_

(Ranks are from 1 to 60 with 60 being the highest rank).

Appendix I

Instructor Report For  
Kansas State Evaluation

## EXPERIMENTAL TEACHER/COURSE EVALUATION REPORT

PREPARED FOR:

MARCH, 1975

ITEM NUMBER	NUMBER OMITTED	NUMBER TRUE	NUMBER FALSE	AVERAGE RESPONSE*
1	0	24	0	1.00
2	0	24	0	1.00
3	0	24	0	1.00
4	0	3	21	1.88
5	0	24	0	2.00
6	0	24	0	2.00
7	0	24	1	1.96
8	0	22	2	1.08
9	0	0	24	1.00
10	0	24	0	1.00
11	0	2	22	1.92
12	0	23	1	1.96
13	0	0	24	1.00
14	0	22	2	1.08
15	1	21	2	1.91
16	0	23	1	1.04
17	0	1	23	1.04
18	0	22	2	1.92
19	0	23	1	1.04
20	0	21	3	1.13
21	0	22	2	1.08
22	0	23	1	1.04
23	0	2	22	1.08
24	0	23	1	1.96
25	0	1	23	1.04
26	1	12	11	1.52
27	0	24	0	1.00
28	0	21	3	1.88
29	0	3	21	1.88
30	0	22	2	1.08
31	0	24	0	2.00
32	0	3	21	1.88
33	0	2	22	1.92
34	0	23	1	1.04
35	0	0	24	2.00
36	0	1	23	1.96
37	0	24	0	1.00
38	0	24	0	1.00
39	0	24	0	1.00
40	0	1	23	1.96
41	0	16	8	1.33
42	0	21	3	1.13
43	0	23	1	1.04
44	0	23	1	1.04
45	0	8	16	1.33
46	0	1	23	1.96
47	0	5	19	1.21
48	0	24	0	1.00
49	0	15	9	1.63
50	0	2	22	1.92
51	0	3	21	1.88
52	0	23	1	1.96
53	1	22	1	1.04
54	0	1	23	1.04
55	0	4	20	1.83
56	0	23	1	1.04
57	0	21	3	1.13
58	0	1	23	1.04

\*RESPONSES TO NEGATIVELY WORDED ITEMS REVERSED FOR COMPUTING AVERAGES.

RESPONSES FOR 10-POINT TEACHER OVERALL RATING SCALE

(LOWEST RATING) 1 2 3 4 5 6 7 8 9 10 (HIGHEST RATING)

NUMBER OF RESPONSES 0 0 0 1 0 0 0 1 13 9

NUMBER OF OMISSIONS = 0

AVERAGE RATING = 9.13

TOTAL NUMBER OF STUDENTS PARTICIPATING = 24

## VITA

Edith Houston Carter was born in Charlotte, North Carolina, on October 12, 1936. She graduated from East Mecklenburg High School in Charlotte, North Carolina, in 1955, received a Bachelor of Science degree in elementary education with a minor in social studies from Appalachian State University in 1959 and a Master of Arts degree in social studies and education from Appalachian State University in 1960.

The author served as a transcript analyst for the Florida State Department of Education from 1961 to 1965 while her husband, Fletcher, was completing the requirements for the Ph.D. in the Department of Higher Education at Florida State University.

The author was employed by New River Community College in 1971 as a part time instructor in History. In 1972 she was given the responsibility for establishing and developing an Office of Institutional Research. Since 1974 she has served as Director of Institutional Research for the college.

The author has been active in music groups and in civic and professional organizations. She is a member of the Association for Institutional Research, the American Educational Research Association, the Virginia Educational Research Association. She serves as Vice Chairman and Newsletter Editor for the Southern Association for Community College Research, a division of A.E.R.A.

THE USE OF SUMMATED RATINGS  
IN FACULTY EVALUATION

by

Edith Houston Carter

(ABSTRACT)

Student evaluation of instruction was investigated using summated ratings obtained from three different types of evaluation instruments: (1) a standardized form developed by Kansas State University (2) a single item form on which the student indicated an overall rating of the instructor and (3) a form on which students constructed their own items and rated the instructor.

The study attempted to answer the following questions: (1) To what extent can summated ratings distinguish among instructors? (2) How strongly do summated ratings based on student constructed items correlate with ratings from other evaluation instruments? (3) What item topics identified by students are common to all or nearly all instructors? (4) Do profiles of instructors based on summated ratings provide a basis for distinguishing among instructors on the basis of certain personal characteristics?

The study utilized students and faculty from three different types of educational institutions: A community college, a senior college and a state university.

Summated scores from individual students were used as a primary basis for analysis. Means of summated scores from the three types of instruments and subsets of items identified from the student constructed items form were analyzed for each instructor using a one-way analysis of variance. Duncan's New Multiple Test was used to isolate groups without significant mean differences for each measure of evaluation. Mean ranks for each instructor were used to obtain correlation coefficients between each measure of evaluation. Kendall's coefficient of concordance was used to determine the degree of agreement between the measures of evaluation. Rankings were used in a pattern analysis (Johnson's MAX procedure) to determine whether rank profiles could be related to personal and professional characteristics of the instructor.

Findings from the study indicated that for the vast majority of faculty members, practical differences could not be determined on the basis of summated ratings. An alternative method based on ranks of sub-groups identified by the New Multiple Range Test provides a more practical approach for distinguishing among faculty members.

It was concluded, from the moderate to high correlations between the rankings for each measure of evaluation, that the instructor would receive basically the same ranking no matter which instrument was used for evaluation. High values obtained from the Kendall

coefficient of concordance indicated a high degree of association between the measures of evaluation.

Five item topics were identified from the student constructed items as being common to all or nearly all instructors. These items defined three areas of concern that the student has for instruction: (1) the instructor knows the subject, (2) the subject is presented well, and (3) the instructor relates positively to the student personally and professionally.

Instructors in courses in which immediate application of course work is evident will receive higher ratings than instructors in those courses in which content application is difficult to discern, such as English and history. The teaching specialty of the instructor, therefore, does make a difference in the kind of evaluation the instructor receives.