1970

THE MEASUREMENT OF THE INSTRUCTOR VARIABLE IN THE INSTRUCTIONAL PROCESS BY A RANK ORDER, FORCED CHOICE PROCEDURE ALONG FIVE BASIC DIMENSIONS

Martha Louise Green

THE MEASUREMENT OF THE INSTRUCTOR VARIABLE IN THE INSTRUCTIONAL PROCESS BY A RANK ORDER, FORCED CHOICE PROCEDURE ALONG FIVE BASIC DIMENSIONS

Martha Louise Green

A Thesis

presented to

the Faculty of the

Department of Psychology

of

Virginia Commonwealth University

in Partial Fulfillment

of the Requirements for the Degree of

Master of Science

July, 1979

THESIS COMMITTEE

CHAIRMAN

Joel R. Butler, Ph. D.
Professor of Psychology

MEMBERS

Michael W. Etkin, Ph. D.
Assistant Professor
of Psychology

Charles D. Noblin, Ph. D.
Professor of Psychology

APPROVED

[redacted]
Joel R. Butler, Ph. D.

[redacted]
Michael W. Etkin, Ph. D.

[redacted]
Charles D. Noblin, Ph. D.

ACKNOWLEDGMENTS

The writer wishes to express her deep appreciation and great thanks to the members of her thesis committee -- Dr. Michael Etkin, Dr. Charles D. Noblin, and especially to a most understanding and tolerant Chairman, Dr. Joel R. Butler -- for their time, guidance, and interest. A particular expression of gratitude is acknowledged to Dr. Harry Anderson for his advise in the treatment and final analysis of the data. To Dr. Charles L. Darby, Mr. Neil Walker, Bette L. Jones, and Dr. Glenn Ware of the University of Georgia, special thanks and deep appreciation are due for the actual computer analysis of the data. The writer is grateful to Webb B. Blackman, Ann Revere Bristow, Patricia Arleen Brooks, Susan Anita Cleary, Susanne Durling, Helen Elizabeth Hare, Marion Elizabeth MacCorkle, and Barbara Gail Sabin for their valuable assistance in the compilation of the data. Special acknowledgements are made to Alice Mary Morrell and Catherine Louise Martin for their great efforts in the preparation of the drafts and the final thesis.

Most of all, the writer is grateful to her parents, Mr. and Mrs. Herman Patrick Green, for their constant support, patience, generous encouragement, and especially for their love.

Appreciation is also extended to Dr. Edwin R. Thomas, Chairman, Department of Psychology, Virginia Commonwealth University, for his guidance, patience and friendship throughout the years of my graduate training.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Student evaluations of college professors is a phenomenon that dates back to the Middle Ages at the University of Padua in Italy. At that time, the students hired their own professors, basing their selections on the professors' earned reputations (Werdell, 1967).

The tradition of student evaluation of professors in America has been somewhat different. Here, the students have not been granted a voice in the educational system. Their evaluations have been viewed by faculty members as a method of "letting off steam." In 1924, the phenomenon of student evaluation, as it is known today, began on the Harvard University campus. Harvard students published a Confidential Guide to Courses (Eble, 1970) which gave a review of students' ratings of courses, professors, examinations, etc. Today, this publication is still doing a thriving business. At about the same time, the University of Washington began an evaluation program, and in 1954 the University of Michigan launched a professor evaluation program (Slobin & Nichols, 1969).

By 1949, the idea of students rating their faculty was becoming more acceptable. Most of the large, well-known universities

were beginning to risk giving endorsement of such ratings. These ratings were not sophisticated and many problems were encountered in application and usage. The methods of evaluation used today run the gamut from multiple choice questions to simple ratings -- sometimes on one dimension and sometimes on multiple dimensions.

Student ratings of faculty did not come into vogue until the fall of 1964. Incidents, such as the Berkeley "rebellion" began to stir up student cries for student participation and faculty involvement in the educational community (Falk, 1968). Students' ratings of faculty now began to be recognized as a valuable tool for the academic community.

The fact that students can be "experts" in evaluating instruction has long been a bone of contention among faculty and students (McKeachie, 1969 a). There are many arguments, most of which are not supported by evidence, that the student does not recognize effective and good instruction. Some statements to support this argument are: "Students cannot really evaluate a teacher until they have left college and obtained some perspective on what was really valuable to them" and "Students rate teachers on their personalities -- not on how much they've learned" (McKeachie, 1969 b p. 214).

Opposition to student ratings of teachers seems to be motivated by a distrust of the student and a desire to keep him in his place in the academic environment. The student does have a

vested interest in his classroom experiences (Kent, 1967). Academic freedom is now being extended to students as well as professors. Students probably are the best judges of teaching because they are in the classroom now, and they can make direct observations of the teacher, which is something that faculty peers or department chairmen have not done for years.

Despite the doubt among instructors that students are incompetent to judge teaching, they, the students, are the instructor's primary audience. The instructor addresses his communication to the student. If this communication is unclear to the students, then the instructor has failed to increase their understanding of the course content (Langen, 1966). Students rarely, however, are capable of judging what a course should have been in terms of course content (Renner, 1967), but they can effectively judge his presentation of content (Brogan, 1968).

Slobin and Nichols (1969) have enumerated some frequent objections to student evaluations of teaching and the subsequent lack of evidence for them. One such objection is that student ratings are influenced by variables irrelevant to teaching. But Slobin and Nichols (1969) point out studies which show that such factors as age of student, sex of student and instructor, student's grades, etc. are not correlated with student ratings of instruction. A second objection is that

student ratings reflect only the instructor's personality. This may be true if the rating forms are poorly constructed, but it is possible to construct questionnaires which do indeed tap areas other than personality. A third objection is that students cannot evaluate the goals of teaching. Students are not being asked to set the goals but are asked to evaluate how well the teacher is achieving his goals. A man should be judged by his peers, is a fourth objection. Student evaluations do not violate this. Peers are not expected to be the best judges of teaching ability, but they should be the best judges of the goals of teaching. A fifth objection is that overemphasis on teaching has bad consequences. This could be an objection only if good teaching is not essential. Slobin and Nichols (1969) quoted E. R. Guthrie as saying, "It is well to remember that student evaluation is continuous and inescapable. The only question is whether or not we care to know what it is."

Developing scales for student evaluation of instruction began about twenty years ago in a real systematic manner with the Purdue Rating Scale (Remmers & Elliott, 1950). This is a graphic ten-point rating scale consisting of ten qualities of a teacher. The scale can be used to develop profiles for each faculty member against norms that have been developed.

In the spring of 1961, on the campus of Grinnell College, Isaacson, McKeachie, Milholland, et al (1963) factor analyzed

five factors of instruction from student evaluations: skill, overload, structure, student rapport, and group interaction. With further study, these same men (1964), found six reliable factors: skills, overload (stability), structure, feedback, group interaction, student-teacher rapport. Cosgrove (1959) found four specific factors of instruction that he called the following: knowledge and organization of subject matter, adequacy of relations with students in class, adequacy of plans and procedures in class, and enthusiasm in working with students. Coffman (1954), on a five point scale, factor analyzed eighteen factors into four: empathy, organization, personal appearance, and verbal fluency. Hoffman (1963) found seven attributes that students saw in the excellent teacher at Hofstra College. They are as follows: attitude toward students (treats them as individuals), presentation, general worthiness as a person, knowledge of subject matter, stimulation of thought and interest, professor's attitude toward teaching (dedication) and tests and grading. Morton (1963) found seven factors that students said they expected to find in a good teacher. They were: knowledge; attractive, active and interesting human being; fellowship (find out through the culture and personality of the teacher a key to why things are important); listener and friend, as well as a lecturer and critic; leader who follows the master; and a teacher who empathizes.

From the above studies and others, five factors were found which appear consistent across all studies in instructional evaluation: skill-ability-presentation, stability, organization, feedback, and instructional image. These are the five factors to be used in this study of instructional evaluation.

Rating scales used in psychological studies began with Galton and were first used in psychophysical experiments specifically in scaling stimuli (Baker & Butler, 1960). Cattell, however, has been given credit for the origin of rank order scaling (Guilford, 1936). With a rating by rank order technique each faculty member is compared to every other faculty member on the same five dimensions, and students are forced to make a decision between instructors.

A chief claim of the forced choice rating scale is that it reduces a deliberate faking of scores by students who wish to assign a very high or very low rating to a teacher regardless of the objectivity of the rating. Using the forced choice rating scale, one is able to fake better than chance -- but not that much better (Lowell & Haner, 1955). Rating scales were found to be of great interest in industry, especially in the area of job evaluation (Baker, 1970). It is in this manner of evaluating performance on the job that rating scales will be used in the present study. The effectiveness of the instructor as a variable in the learning process is so complex a phenomenon that only something as sensitive as another human observer can report these characteristics.

Rating scales are found to be the most useful methods in achieving this end.

Rating scales must have criteria that make it possible to use them as measuring devices. Remmers in the Handbook of Research on Teaching (1963) suggests five criteria:

1. Objectivity -- instrument should yield verifiable, reproducible data
2. Reliability -- should yield same values under same conditions
3. Sensitivity -- should yield fine distinctions such as those made in communicating the objectives of investigation
4. Validity -- definitional, construct, concurrent, and predictive
5. Utility -- should be efficient and practical (p. 330).

The purpose of this study was a rating of professors as a variable in the learning process through a dimension rank order technique. It was hypothesized that a rank ordering of instructors would yield meaning in standard scores along five main dimensions of instruction by a paired comparison transformation, and that significant differences could be demonstrated between instructors on all five dimensions of instruction. It was also hypothesized that there would be rater agreement on the rank order of an instructor on each dimension.

CHAPTER II

METHOD

Subjects

The subjects were 316 undergraduate students enrolled in day and evening psychology courses, above the introductory level, for a full semester at Virginia Commonwealth University.

Material

The material for this study was a teacher evaluation instrument, which consisted of five dimensions of instruction acquired from previous studies. These dimensions were ranked by students at Virginia Commonwealth University, and then comments were analyzed to form five basic dimensions of instruction (see Appendix 1). The initial five dimensions of instruction were restated in the student's own words, and dealt with knowledge, presentation, work load, tests, and effective interaction with students. The names of the full-time members in the Psychology Department of Virginia Commonwealth University were arranged on five lists, the order of which was obtained by random numbering. Each of the five dimensions, along with a randomly ordered list of faculty names, a set of standard instructions, and a cover sheet requesting information such as name of professor, class, and student's academic year, composed the instrument (see Appendix 2).

Procedure

The instrument was administered by students to students in psychology classes from April 6, 1970 to April 15, 1970. During class sessions, the professor was asked to leave the room and the students were given the instrument. A set of standard instructions (which was also attached to the rating sheets) was read to the students. After the evaluation was completed, the data were collected by the student research evaluators, and the professor returned to his classroom. The average period of time involved per class in the collection of data was fifteen minutes.

The data were transferred to IBM quality control cards according to a format given in a statistical procedure incorporated in the Dykstra Analysis Program (Dykstra, 1970).

CHAPTER III

RESULTS AND DISCUSSION

When the data in the study were submitted to the Dykstra Analysis (1970) (see Appendix 3) based on the Bradley-Terry method of paired comparisons (1955), it revealed clear evidence that students could reliably discriminate between instructors on various dimensions of instruction by a rank order forced choice technique. In addition, these choices could significantly differentiate between instructors with the data revealing consistent differential preferences (from highest to lowest) on a given dimension. The range limits among ranks revealed no significant differences occurring between instructors. But, these same upper-lower limits marked a significant difference between the ranks represented by the limits of a given range. For example, no difference was revealed on Dimension Two between ranks 2 and 3, or between 3 and 4, but a difference was shown between ranks 2 and 4.

A basic tenet upon which the method of paired comparisons rests is the probability that one value (instructor) will always be ranked superior to another. This is the method of quantification, and the binomial model is to be followed. In this instance, the test of significance used is a between groups test of goodness of fit, which is a test of agreement that permits the students to differ in their judgments of instructors (Bradley & Terry, 1954a).

In subjective testing, such as rank ordering of instructors, an assumption can be made, a priori, that the standards of judging vary by students, time, conditions, etc., or any combination of these. However, this study has shown that tests of instructor differences may be performed, and a measure of agreement among students obtained.

In the quantification of paired comparisons, the first problem, in this instance, is to determine the actual number of comparisons between instructors (from their respective rank orders) and to find the differences in terms of higher and lower rankings between any two given instructors. Results of these comparisons can be seen in Table 1.

It may be seen, for example, that when instructor 1 is compared to instructor 2, that instructor 2 was rated superior six times but that instructor 1 was rated superior forty-four times. This initial measure of discrepancy shows the actual numerical difference, but does not reveal the explicit agreement or lack of agreement between raters; nor does it show the combinatorial (rather than linear variate) results upon which further tests of differences are based.

When all preferences are combined for each dimension individually (see Tables 2, 3, 4, 5, 6), the number of times a given instructor was rated higher (won) or lower (lost) and the total number of times that he was compared to other instructors are given. For example, for Dimension One, it may be seen that instructor 1 won 247 times and lost 137 times from a total number of 411 comparisons -- for about a 2 to 1 won-lost record, while instructor 2 was rated higher 216 times and rated lower 743 times from a total of 959 comparisons -- for about a 1 to 3-1/2 won-lost record.

Table 1

Paired Comparisons Between All Instructors on Dimension One

Lower Instructor	Higher Instructor	Times Higher	Lower Instructor	Higher Instructor	Times Higher
1	1	0.0	3	1	4.0
1	2	6.0	3	2	13.0
1	3	0.0	3	3	0.0
1	4	8.0	3	4	3.0
1	5	20.0	3	5	3.0
1	6	0.0	3	6	0.0
1	7	12.0	3	7	3.0
1	8	0.0	3	8	4.0
1	9	1.0	3	9	2.0
1	10	5.0	3	10	0.0
1	11	6.0	3	11	6.0
1	12	7.0	3	12	0.0
1	13	13.0	3	13	11.0
1	14	4.0	3	14	0.0
1	15	13.0	3	15	9.0
1	16	11.0	3	16	13.0
1	17	2.0	3	17	0.0
1	18	19.0	3	18	13.0
1	19	5.0	3	19	0.0
1	20	5.0	3	20	4.0
2	1	44.0	4	1	7.0
2	2	0.0	4	2	2.0
2	3	5.0	4	3	0.0
2	4	40.0	4	4	0.0
2	5	64.0	4	5	7.0
2	6	23.0	4	6	0.0
2	7	51.0	4	7	2.0
2	8	11.0	4	8	2.0
2	9	55.0	4	9	1.0
2	10	17.0	4	10	3.0
2	11	45.0	4	11	3.0
2	12	11.0	4	12	0.0
2	13	56.0	4	13	4.0
2	14	22.0	4	14	3.0
2	15	72.0	4	15	3.0
2	16	61.0	4	16	1.0
2	17	6.0	4	17	2.0
2	18	93.0	4	18	10.0
2	19	24.0	4	19	0.0
2	20	43.0	4	20	5.0

Lower Instructor	Higher Instructor	Times Higher	Lower Instructor	Higher Instructor	Times Higher
5	1	11.0	7	1	12.0
5	2	4.0	7	2	20.0
5	3	0.0	7	3	1.0
5	4	6.0	7	4	15.0
5	5	0.0	7	5	33.0
5	6	2.0	7	6	5.0
5	7	12.0	7	7	0.0
5	8	3.0	7	8	4.0
5	9	8.0	7	9	12.0
5	10	6.0	7	10	4.0
5	11	7.0	7	11	19.0
5	12	3.0	7	12	3.0
5	13	13.0	7	13	22.0
5	14	3.0	7	14	0.0
5	15	14.0	7	15	16.0
5	16	9.0	7	16	13.0
5	17	0.0	7	17	2.0
5	18	28.0	7	18	20.0
5	19	4.0	7	19	3.0
5	20	1.0	7	20	13.0
6	1	17.0	8	1	5.0
6	2	10.0	8	2	3.0
6	3	2.0	8	3	1.0
6	4	14.0	8	4	2.0
6	5	18.0	8	5	7.0
6	6	0.0	8	6	4.0
6	7	18.0	8	7	2.0
6	8	5.0	8	8	0.0
6	9	11.0	8	9	2.0
6	10	13.0	8	10	1.0
6	11	10.0	8	11	5.0
6	12	9.0	8	12	1.0
6	13	16.0	8	13	1.0
6	14	0.0	8	14	2.0
6	15	21.0	8	15	7.0
6	16	7.0	8	16	2.0
6	17	0.0	8	17	0.0
6	18	32.0	8	18	8.0
6	19	10.0	8	19	5.0
6	20	5.0	8	20	2.0

Lower Instructor	Higher Instructor	Times Higher	Lower Instructor	Higher Instructor	Times Higher
9	1	39.0	11	1	13.0
9	2	31.0	11	2	13.0
9	3	3.0	11	3	2.0
9	4	17.0	11	4	12.0
9	5	46.0	11	5	14.0
9	6	16.0	11	6	0.0
9	7	36.0	11	7	3.0
9	8	5.0	11	8	5.0
9	9	0.0	11	9	12.0
9	10	11.0	11	10	2.0
9	11	21.0	11	11	0.0
9	12	9.0	11	12	1.0
9	13	38.0	11	13	7.0
9	14	13.0	11	14	5.0
9	15	36.0	11	15	10.0
9	16	30.0	11	16	14.0
9	17	0.0	11	17	0.0
9	18	51.0	11	18	20.0
9	19	20.0	11	19	1.0
9	20	23.0	11	20	5.0
10	1	6.0	12	1	0.0
10	2	4.0	12	2	2.0
10	3	0.0	12	3	0.0
10	4	3.0	12	4	3.0
10	5	4.0	12	5	0.0
10	6	1.0	12	6	0.0
10	7	9.0	12	7	2.0
10	8	1.0	12	8	0.0
10	9	2.0	12	9	0.0
10	10	0.0	12	10	3.0
10	11	4.0	12	11	4.0
10	12	5.0	12	12	0.0
10	13	3.0	12	13	5.0
10	14	2.0	12	14	0.0
10	15	4.0	12	15	1.0
10	16	2.0	12	16	0.0
10	17	0.0	12	17	0.0
10	18	10.0	12	18	0.0
10	19	1.0	12	19	0.0
10	20	3.0	12	20	0.0

Lower Instructor	Higher Instructor	Times Higher	Lower Instructor	Higher Instructor	Times Higher
13	1	16.0	15	1	22.0
13	2	10.0	15	2	33.0
13	3	0.0	15	3	2.0
13	4	13.0	15	4	17.0
13	5	32.0	15	5	40.0
13	6	4.0	15	6	15.0
13	7	17.0	15	7	27.0
13	8	3.0	15	8	10.0
13	9	2.0	15	9	27.0
13	10	11.0	15	10	14.0
13	11	7.0	15	11	32.0
13	12	3.0	15	12	12.0
13	13	0.0	15	13	34.0
13	14	1.0	15	14	9.0
13	15	18.0	15	15	0.0
13	16	19.0	15	16	42.0
13	17	3.0	15	17	0.0
13	18	32.0	15	18	50.0
13	19	1.0	15	19	12.0
13	20	15.0	15	20	17.0
14	1	3.0	16	1	19.0
14	2	16.0	16	2	14.0
14	3	0.0	16	3	1.0
14	4	13.0	16	4	7.0
14	5	14.0	16	5	29.0
14	6	2.0	16	6	8.0
14	7	7.0	16	7	16.0
14	8	0.0	16	8	9.0
14	9	7.0	16	9	19.0
14	10	0.0	16	10	7.0
14	11	3.0	16	11	25.0
14	12	0.0	16	12	8.0
14	13	14.0	16	13	15.0
14	14	0.0	16	14	9.0
14	15	7.0	16	15	27.0
14	16	12.0	16	16	0.0
14	17	1.0	16	17	0.0
14	18	22.0	16	18	41.0
14	19	2.0	16	19	12.0
14	20	15.0	16	20	13.0

Lower Instructor	Higher Instructor	Times Higher	Lower Instructor	Higher Instructor	Times Higher
17	1	2.0	19	1	15.0
17	2	0.0	19	2	6.0
17	3	0.0	19	3	0.0
17	4	3.0	19	4	15.0
17	5	0.0	19	5	27.0
17	6	0.0	19	6	7.0
17	7	1.0	19	7	17.0
17	8	0.0	19	8	1.0
17	9	0.0	19	9	8.0
17	10	0.0	19	10	10.0
17	11	0.0	19	11	7.0
17	12	0.0	19	12	6.0
17	13	3.0	19	13	22.0
17	14	1.0	19	14	6.0
17	15	3.0	19	15	18.0
17	16	0.0	19	16	12.0
17	17	0.0	19	17	0.0
17	18	2.0	19	18	33.0
17	19	0.0	19	19	0.0
17	20	1.0	19	20	9.0
18	1	32.0	20	1	7.0
18	2	19.0	20	2	10.0
18	3	3.0	20	3	1.0
18	4	36.0	20	4	13.0
18	5	31.0	20	5	23.0
18	6	5.0	20	6	4.0
18	7	23.0	20	7	10.0
18	8	11.0	20	8	2.0
18	9	14.0	20	9	9.0
18	10	12.0	20	10	4.0
18	11	26.0	20	11	10.0
18	12	9.0	20	12	0.0
18	13	29.0	20	13	16.0
18	14	4.0	20	14	11.0
18	15	33.0	20	15	17.0
18	16	38.0	20	16	11.0
18	17	1.0	20	17	2.0
18	18	0.0	20	18	23.0
18	19	6.0	20	19	4.0
18	20	15.0	20	20	0.0

Table 2

Summary of Dimension One

(Wins and Losses Out of Total Comparisons)

Instructor	Wins	Losses	Tries
1	274.0	137.0	411.0
2	216.0	743.0	959.0
3	21.0	88.0	109.0
4	240.0	55.0	295.0
5	414.0	136.0	550.0
6	96.0	218.0	314.0
7	268.0	217.0	485.0
8	76.0	60.0	136.0
9	192.0	445.0	637.0
10	123.0	64.0	187.0
11	240.0	139.0	379.0
12	87.0	20.0	107.0
13	322.0	207.0	529.0
14	95.0	138.0	233.0
15	330.0	416.0	746.0
16	298.0	280.0	578.0
17	19.0	16.0	35.0
18	508.0	348.0	856.0
19	110.0	219.0	329.0
20	194.0	177.0	371.0
Sum	4123.0	4123.0	8246.0

Table 3

Summary of Dimension Two

(Wins and Losses Out of Total Comparisons)

Instructor	Wins	Losses	Tries
1	317.0	97.0	414.0
2	165.0	860.0	1025.0
3	15.0	118.0	133.0
4	265.0	34.0	299.0
5	302.0	258.0	560.0
6	102.0	196.0	298.0
7	326.0	188.0	514.0
8	101.0	53.0	154.0
9	134.0	532.0	666.0
10	112.0	110.0	222.0
11	288.0	123.0	411.0
12	59.0	35.0	94.0
13	343.0	210.0	553.0
14	112.0	138.0	250.0
15	293.0	477.0	770.0
16	391.0	210.0	601.0
17	18.0	20.0	38.0
18	669.0	234.0	903.0
19	98.0	235.0	333.0
20	194.0	176.0	370.0
Sum	4304.0	4304.0	8608.0

Table 4

Summary of Dimension Three

(Wins and Losses Out of Total Comparisons)

Instructor	Wins	Losses	Tries
1	213.0	177.0	390.0
2	336.0	686.0	1022.0
3	27.0	100.0	127.0
4	211.0	94.0	305.0
5	201.0	351.0	552.0
6	90.0	212.0	302.0
7	287.0	224.0	511.0
8	87.0	59.0	146.0
9	240.0	425.0	665.0
10	110.0	107.0	217.0
11	280.0	130.0	410.0
12	73.0	34.0	107.0
13	357.0	172.0	529.0
14	127.0	120.0	247.0
15	401.0	367.0	768.0
16	440.0	159.0	599.0
17	23.0	11.0	34.0
18	491.0	402.0	893.0
19	141.0	206.0	347.0
20	135.0	234.0	369.0
Sum	4270.0	4270.0	8540.0

Table 5

Summary of Dimension Four

(Wins and Losses Out of Total Comparisons)

Instructor	Wins	Losses	Tries
1	265.0	146.0	411.0
2	271.0	738.0	1009.0
3	28.0	97.0	125.0
4	220.0	68.0	288.0
5	305.0	239.0	544.0
6	96.0	198.0	294.0
7	259.0	233.0	492.0
8	65.0	92.0	157.0
9	210.0	438.0	648.0
10	119.0	115.0	234.0
11	262.0	156.0	418.0
12	83.0	24.0	107.0
13	310.0	209.0	519.0
14	161.0	88.0	249.0
15	399.0	370.0	769.0
16	436.0	159.0	595.0
17	16.0	19.0	35.0
18	499.0	379.0	878.0
19	145.0	196.0	341.0
20	106.0	291.0	397.0
Sum	4255.0	4255.0	8510.0

Table 6

Summary of Dimension Five

(Wins and Losses Out of Total Comparisons)

Instructor	Wins	Losses	Tries
1	263.0	126.0	389.0
2	334.0	657.0	991.0
3	56.0	63.0	119.0
4	219.0	72.0	291.0
5	205.0	336.0	541.0
6	79.0	221.0	300.0
7	348.0	144.0	492.0
8	76.0	68.0	144.0
9	169.0	477.0	646.0
10	89.0	139.0	228.0
11	257.0	146.0	403.0
12	76.0	32.0	108.0
13	330.0	178.0	508.0
14	139.0	119.0	258.0
15	323.0	416.0	739.0
16	429.0	140.0	569.0
17	28.0	15.0	43.0
18	483.0	383.0	866.0
19	161.0	158.0	319.0
20	105.0	279.0	384.0
Sum	4169.0	4169.0	8338.0

In the interest of accuracy and in the testing of the null hypothesis, which states that there is no significant difference between instructors, a compensation for the unequal repetitions on pairs and the unequal blocks needs to be made so that the total final percent preference will equal a total of one (Bradley, 1955). A repetition is defined as a single set of incomplete blocks or cells of size two with pairs of instructors appearing together just once $\binom{t}{2}$ (Bradley, 1954b), and t is used to cite the number of instructors. This correction can be accomplished through an iterative procedure (Dykstra, 1970). Maximum likelihood estimators of the parameters must be made and specified as $P_1 \cdots\cdots P_t$. The ratio P_i/P_j measures the relative frequency of the occurrence of rank one for instructor i as compared with instructor j for the particular paired comparison in question (Bradley & Terry, 1952). Specification of symbols for the iterative formula are as follows:

$$P_i = \frac{a_i}{n\,(t-1)^2 - a_i\,(t-2)} \qquad \text{or} \qquad P_i = \frac{a_i \,/\, n}{\sum_{j \neq i} 1/\,(P_i + P_j)}$$

(Bradley, 1966) (Bradley, 1956)

P_i = maximum likelihood estimator of preference of instructor i

$a_i = 2_n\,(t-1) - \sum_{j \neq i} \sum_{k=1}^{n} r_{ijk}$

r_{ijk} is the rank of instructor i in the comparison with instructor j in the kth of n repetitions of the paired comparisons (Bradley, 1955).

t = number of instructors

n = number of times instructor i and instructor j were compared

The iterative formula used in this study (Dykstra, 1970) is:

$$P_i = A_i / \sum_{j \neq i} (N_{ij}/(P_i+P_j))$$

The first estimators of the parameters (ratings)are substituted into the right side of the formula, the next estimators are substituted, etc. until the equalities hold. Iterations continue, until the largest change in any P_i is less than $0.004/T$ or until the series of iterations begin to diverge. For example, on Dimension One, forty iterations were needed before the series of iterations began to diverge.

A test must then be run to see if all P_i are equal. This test is a X^2 (chi square) statistic with t-2 df (Dykstra, 1970).

$$X^2 = 2((\sum A_i) \text{ LN2-}B_1)).$$

The statistic B_1 is used for a test of instructor equality assuming homogeneous repetitions of paired comparisons.

$$B_1 = n \sum_{i<j} \log (P_i+P_j) - \sum_i (2n(t-1) - \sum r_{ij}) \log P_i$$

The statistic B_1 is required for the combined test of instructor equality, test for homogeneous repetitions, and analysis of paired comparisons with the instructors in factorial arrangements (Bradley, 1954 b). The formula for B_1 is modified for the Dykstra analysis and occurs as:

$$B_1 = \sum_{i<j} N_{ij} \text{ LN}(P_i+P_j) - \sum_i A_i \text{LN} P_i$$

Having completed the chi square analysis, a standard deviation must be computed. The standard deviation used is given by:

$$\sigma = 50 / \sqrt{T/(T-1)/2}$$

where T is the total number of comparisons. The power of the test is based on Tang's tables where, if the standard deviation is ≥ 1 but < 2 the comparison between two instructors is significant at the .05 level, and if the standard deviation is > 2 then the comparison between two instructors is significant at the .01 level (Bradley, 1955).

In considering the ratings of instructors, it is assumed that these instructors have true ratings or preferences assigned to them by the student raters. These instructor ratings or parameters are designated by the symbol π_1 ---- π_t, and are on a subjective continuum which is specified by $\pi_i \geq 0$ and $\sum_i \pi_i = 1$ (Bradley, 1955). Table 7 first shows, for each instructor, the initial percent preference, which is the raw data before it has been corrected for unequal repetitions on pairs. For example, instructor 1 has an initial percent preference of 0.095238. Then, in the final percent preference column, the data has been corrected for unequal repetitions on pairs and, therefore, each preference for an instructor is noted in proportions and can then be ranked. Instructor 1 then has a final percent preference of 0.069305. As specified, the sum of the final percent preferences equals one. The final percent preferences are then ranked, with a rank of 1 being the instructor with the highest final percent preference; and a rank of 20, the instructor with the lowest. In the case of dimension one, instructor 12 is ranked first, and instructor 3 is ranked 20. This can be further shown for each instructor on each dimension (see Tables 8, 9, 10, 11).

Table 7

Initial and Final Percent Preference
Score and Rank Order of Dimension One

Instructor	Initial Percent Preference	Final Precent Preference	Rank
1	0.095238	0.069305	4
2	0.015070	0.009764	19
3	0.012404	0.006616	20
4	0.186770	0.149441	2
5	0.138092	0.100587	3
6	0.022652	0.014066	17
7	0.061034	0.039913	11
8	0.062500	0.040377	10
9	0.022204	0.013447	18
10	0.091860	0.068821	5
11	0.083304	0.055260	6
12	0.186295	0.153951	1
13	0.075676	0.052008	7
14	0.034965	0.021258	15
15	0.040078	0.024827	14
16	0.053044	0.033060	13
17	0.058824	0.048043	9
18	0.071348	0.048534	8
19	0.025755	0.016866	16
20	0.054540	$\frac{0.033858}{1.000000}$ (within .004)	12

Table 8

Initial and Final Percent Preference
Score and Rank Order of Dimension Two

Instructor	Initial Percent Preference	Final Percent Preference	Rank
1	0.146759	0.103736	2
2	0.009997	0.004574	19
3	0.006646	0.002345	20
4	0.290889	0.293643	1
5	0.058032	0.029992	11
6	0.026660	0.011869	16
7	0.083633	0.045988	7
8	0.091155	0.060713	5
9	0.013083	0.005374	18
10	0.050863	0.028410	13
11	0.109714	0.072149	4
12	0.081492	0.044392	9
13	0.079160	0.045874	8
14	0.040966	0.020461	14
15	0.031317	0.014443	15
16	0.089249	0.052239	6
17	0.045226	0.034506	10
18	0.130792	0.090770	3
19	0.021477	0.010062	17
20	0.054833	0.028465	12
		= 1.000000	
		(within .004)	

Table 9

Initial and Final Percent Preference
Score and Rank Order of Dimension Three

Instructor	Initial Percent Preference	Final Percent Preference	Rank
1	0.059564	0.045998	10
2	0.025131	0.020192	18
3	0.014011	0.010813	20
4	0.105658	0.088895	3
5	0.029258	0.021664	16
6	0.021855	0.016547	19
7	0.063174	0.046866	9
8	0.072020	0.058830	7
9	0.028863	0.021437	17
10	0.051330	0.039056	13
11	0.101818	0.084787	5
12	0.101530	0.087117	4
13	0.098483	0.077202	6
14	0.052763	0.040448	12
15	0.054380	0.042203	11
16	0.127131	0.103820	1
17	0.099138	0.096846	2
18	0.060401	0.048923	8
19	0.034772	0.026395	14
20	0.029470	$\frac{0.021963}{1.000000}$ (within .004)	15

Table 10

Initial and Final Percent Preference
Score and Rank Order of Dimension Four

Instructor	Initial Percent Preference	Final Percent Preference	Rank
1	0.087200	0.069958	4
2	0.018960	0.014578	18
3	0.014965	0.010235	20
4	0.145503	0.121526	2
5	0.062938	0.046823	9
6	0.024883	0.018146	16
7	0.055271	0.039156	11
8	0.035852	0.026019	15
9	0.024613	0.017744	17
10	0.051649	0.039092	12
11	0.081215	0.064264	6
12	0.153989	0.141084	1
13	0.072413	0.055112	7
14	0.087834	0.068036	5
15	0.053708	0.039803	10
16	0.126121	0.102214	3
17	0.042440	0.032355	13
18	0.064805	0.051865	8
19	0.037477	0.029009	14
20	0.018811	0.012985	19
		= 1.000000	
		(within .004)	

Table 11

Initial and Final Percent Preference
Score and Rank Order of Dimension Five

Instructor	Initial Percent Preference	Final Percent Preference	Rank
1	0.098984	0.075087	6
2	0.026059	0.018450	17
3	0.044693	0.034493	13
4	0.137996	0.112859	1
5	0.031112	0.020350	16
6	0.018467	0.011963	19
7	0.112840	0.080908	5
8	0.055556	0.040660	10
9	0.018306	0.011953	20
10	0.032601	0.021761	15
11	0.084790	0.064980	7
12	0.111111	0.088994	3
13	0.088901	0.064687	8
14	0.057917	0.039785	11
15	0.039261	0.026766	14
16	0.138880	0.107864	2
17	0.089457	0.083716	4
18	0.062242	0.046587	9
19	0.050901	0.035759	12
20	0.019423	$\frac{0.012379}{1.000000}$ (within .004)	18

When instructor i appears with instructor j in a particular cell or comparison then the probability that instructor i obtains the higher rating (that of rank 1) is $\frac{\pi_i}{(\pi_i + \pi_j)}$ otherwise known as a pairwise preference (Bradley, 1955). This follows the binomial model and rests on the principle of probability. The probability statement is expressed as:

$$P(X_i > X_j) = \frac{\pi_i}{(\pi_i + \pi_j)}$$ (Dykstra, 1960).

The formula for the pairwise preferences used in this study is:

$$PP_{ij} = \frac{P_i}{\frac{(P_i+P_j)-.5}{2}} \times 100$$

$$= \frac{P_i}{(P_i+P_j)} \times 50\text{-}25$$

$$= \frac{P_i}{(P_i+P_j)} \times 25$$

(Dykstra, 1970)

In dimension one (see Table 12) instructor 1 is compared with instructor 2, and instructor 1 was preferred or was rated superior by approximately six standard deviations above instructor 2. This is significant at the .01 level. Significant differences can more readily be noted for each instructor on each dimension in Tables 13, 14, 15, and 16. It can be seen (see Table 13) that instructor 12 was ranked number 1, instructor 4 was ranked number 2, instructor 5 was

Table 12

Pairwise Preferences
Dimension One

Instructor (I)	Instructor (J)	Preference (I,J)	Instructor (I)	Instructor (J)	Preference (I,J)
1	2	18.8255*	3	1	-20.6429*
1	3	20.6429*	3	2	-4.8053*
1	4	-9.1586*	3	4	-22.8803*
1	5	-4.6032*	3	5	-21.9143*
1	6	16.5640*	3	6	-9.0060*
1	7	6.7278*	3	7	-17.8906*
1	8	6.5935*	3	8	-17.9608*
1	9	16.8749*	3	9	-8.5126*
1	10	0.0875	3	10	-20.6150*
1	11	2.8187	3	11	-19.6540*
1	12	-9.4786*	3	12	-22.9398*
1	13	3.5645*	3	13	-19.3574*
1	14	13.2635*	3	14	-13.1325*
1	15	11.8125*	3	15	-14.4798*
1	16	8.8520*	3	16	-16.6625*
1	17	4.5296*	3	17	-18.9481*
1	18	4.4067*	3	18	-19.0019*
1	19	15.2138*	3	19	-10.9127*
1	20	8.5900*	3	20	-16.8271*
2	1	-18.8255*	4	1	9.1586*
2	3	4.8052*	4	2	21.9335*
2	4	-21.9334*	4	3	22.8803*
2	5	-20.5758*	4	5	4.8849*
2	6	-4.5132*	4	6	20.6986*
2	7	-15.1723*	4	7	14.4608*
2	8	-15.2633*	4	8	14.3643*
2	9	-3.9670*	4	9	20.8722*
2	10	-18.7875*	4	10	9.2343*
2	11	-17.4919*	4	11	11.5022*
2	12	-22.0179*	4	12	-0.3716
2	13	-17.0966*	4	13	12.0916*
2	14	-9.2624*	4	14	18.7733*
2	15	-10.8865*	4	15	17.8767*
2	16	-13.5995*	4	16	15.9426*
2	17	-16.5545*	4	17	12.8362*
2	18	-16.6255*	4	18	12.7425*
2	19	-6.6668*	4	19	19.9294*
2	20	-13.8083*	4	20	15.7642*

*significance at or beyond the .05 level if the value of P(I,J) is greater than or equal to 3.3942.

Instructor (I)	Instructor (J)	Preference (I,J)	Instructor (I)	Instructor (J)	Preference (I,J)
5	1	4.6032*	7	1	-6.7278*
5	2	20.5758*	7	2	15.1723*
5	3	21.9143*	7	3	17.8906*
5	4	-4.8849*	7	4	-14.4608*
5	6	18.8657*	7	5	-10.7961*
5	7	10.7961*	7	6	11.9706*
5	8	10.6782*	7	8	-0.1446
5	9	19.1038*	7	9	12.3994*
5	10	4.6877*	7	10	-6.6466*
5	11	7.2710*	7	11	-4.0315*
5	12	-5.2413*	7	12	-14.7060*
5	13	7.9588*	7	13	-3.2895
5	14	16.2767*	7	14	7.6242*
5	15	15.1018*	7	15	5.8253*
5	16	12.6317*	7	16	2.3479
5	17	8.8379*	7	17	-2.3110
5	18	8.7267*	7	18	-2.4367
5	19	17.8202*	7	19	10.1478*
5	20	12.4082*	7	20	2.0518
6	1	-16.5640*	8	1	-6.5935*
6	2	4.5132*	8	2	15.2633*
6	3	9.0060*	8	3	17.9608*
6	4	-20.6986*	8	4	-14.3643*
6	5	-18.8657*	8	5	-10.6782*
6	7	-11.9706*	8	6	12.0817*
6	8	-12.0817*	8	7	0.1446
6	9	0.5623	8	9	12.5081*
6	10	-16.5148*	8	10	-6.5120*
6	11	-14.8550*	8	11	-3.8905*
6	12	-20.8140*	8	12	-14.6111*
6	13	-14.3557*	8	13	-3.1474
6	14	-5.0896*	8	14	7.7551*
6	15	-6.9170*	8	15	5.9619*
6	16	-10.0758*	8	16	2.4911
6	17	-13.6762*	8	17	-2.1675
6	18	-13.7649*	8	18	-2.2934
6	19	-2.2625	8	19	10.2683*
6	20	-10.3245*	8	20	2.1954

*significance at or beyond the .05 level if the value of P(I,J) is greater than or equal to 3.3942.

Instructor (I)	Instructor (J)	Preference (I,J)	Instructor (I)	Instructor (J)	Preference (I,J)
9	1	-16.8748*	11	1	-2.8187
9	2	3.9670*	11	2	17.4919*
9	3	8.5126*	11	3	19.6540*
9	4	-20.8722*	11	4	-11.5022*
9	5	-19.1037*	11	5	-7.2710*
9	6	-0.5623	11	6	14.8550*
9	7	-12.3994*	11	7	4.0315*
9	8	-12.5081*	11	8	3.8905*
9	10	-16.8271*	11	9	15.2140*
9	11	-15.2140*	11	10	-2.7323
9	12	-20.9834*	11	12	-11.7932*
9	13	-14.7277*	11	13	0.7580
9	14	-5.6262*	11	14	11.1093*
9	15	-7.4331*	11	15	9.4998*
9	16	-10.5425*	11	16	6.2842*
9	17	-14.0655*	11	17	1.7465
9	18	-14.1519*	11	18	1.6202
9	19	-2.8190	11	19	13.3082*
9	20	-10.7866*	11	20	6.0038*
10	1	-0.0875	12	1	9.4786*
10	2	18.7875*	12	2	22.0179*
10	3	20.6150*	12	3	22.9398*
10	4	-9.2343*	12	4	0.3716
10	5	-4.6877*	12	5	5.2413*
10	6	16.5148*	12	6	20.8140*
10	7	6.6466*	12	7	14.7059*
10	8	6.5120*	12	8	14.6111*
10	9	16.8271*	12	9	20.9834*
10	11	2.7322	12	10	9.5534*
10	12	-9.5535*	12	11	11.7932*
10	13	3.4787*	12	13	12.3742*
10	14	13.2005*	12	14	18.9336*
10	15	11.7443*	12	15	18.0564*
10	16	8.7754*	12	16	16.1610*
10	17	4.4448*	12	17	13.1077*
10	18	4.3219*	12	18	13.0155*
10	19	15.1586*	12	19	20.0632*
10	20	8.5127*	12	20	15.9860*

*significance at or beyond the .05 level if the value of P(I,J) is greater than or equal to 3.3942.

Instructor (I)	Instructor (J)	Preference (I,J)	Instructor (I)	Instructor (J)	Preference (I,J)
13	1	-3.5645*	15	1	-11.8124*
13	2	17.0966*	15	2	10.8865*
13	3	19.3574*	15	3	14.4798*
13	4	-12.0916*	15	4	-17.8767*
13	5	-7.9588*	15	5	-15.1018*
13	6	14.3557*	15	6	6.9170*
13	7	3.2895	15	7	-5.8253*
13	8	3.1474	15	8	-5.9619*
13	9	14.7278*	15	9	7.4330*
13	10	-3.4787*	15	10	-11.7443*
13	11	-0.7580	15	11	-9.4998*
13	12	-12.3742*	15	12	-18.0563*
13	14	10.4927*	15	13	-8.8437*
13	15	8.8437*	15	14	1.9364
13	16	5.5686*	15	16	-3.5552*
13	17	0.9906	15	17	-7.9647*
13	18	0.8639	15	18	-8.0786*
13	19	12.7561*	15	19	4.7741*
13	20	5.2843*	15	20	-3.8471*
14	1	-13.2635*	16	1	-8.8521*
14	2	9.2624*	16	2	13.5995*
14	3	13.1324*	16	3	16.6625*
14	4	-18.7733*	16	4	-15.9426*
14	5	-16.2767*	16	5	-12.6317*
14	6	5.0896*	16	6	10.0758*
14	7	-7.6242*	16	7	-2.3479
14	8	-7.7551*	16	8	-2.4912
14	9	5.6262*	16	9	10.5425*
14	10	-13.2005*	16	10	-8.7754*
14	11	-11.1093*	16	11	-6.2842*
14	12	-18.9336*	16	12	-16.1610*
14	13	-10.4927*	16	13	-5.5686*
14	15	-1.9365	16	14	5.4318*
14	16	-5.4318*	16	15	3.5552*
14	17	-9.6627*	16	17	-4.6188*
14	18	-9.7705*	16	18	-4.7412*
14	19	2.8802	16	19	8.1091*
14	20	-5.7154*	16	20	-0.2984

*significance at or beyond the .05 level if the value of P(I,J) is greater than or equal to 3.3942.

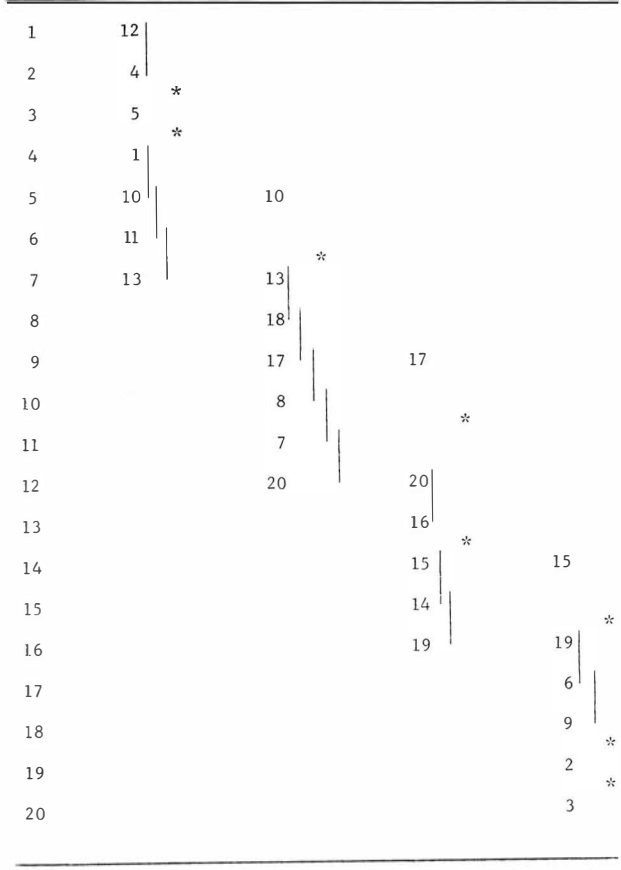
Instructor (I)	Instructor (J)	Preference (I,J)	Instructor (I)	Instructor (J)	Preference (I,J)
17	1	-4.5296*	19	1	-15.2138*
17	2	16.5545*	19	2	6.6668*
17	3	18.9481*	19	3	10.9127*
17	4	-12.8362*	19	4	-19.9294*
17	5	-8.8380*	19	5	-17.8202*
17	6	13.6762*	19	6	2.2625
17	7	2.3110	19	7	-10.1479*
17	8	2.1675	19	8	-10.2683*
17	9	14.0654*	19	9	2.8190
17	10	-4.4449*	19	10	-15.1586*
17	11	-1.7465	19	11	-13.3082*
17	12	-13.1077*	19	12	-20.0632*
17	13	-0.9906	19	13	-12.7561*
17	14	9.6627*	19	14	-2.8802
17	15	7.9647*	19	15	-4.7741*
17	16	4.6188*	19	16	-8.1091*
17	18	-0.1269	19	17	-12.0082*
17	19	12.0083*	19	18	-12.1056*
17	20	4.3300*	19	20	-8.3750*
18	1	-4.4067*	20	1	-8.5900*
18	2	16.6255*	20	2	13.8082*
18	3	19.0019*	20	3	16.8271*
18	4	-12.7425*	20	4	-15.7642*
18	5	-8.7267*	20	5	-12.4082*
18	6	13.7649*	20	6	10.3245*
18	7	2.4367	20	7	-2.0518
18	8	2.2934	20	8	-2.1954
18	9	14.1519*	20	9	10.7866*
18	10	-4.3219*	20	10	-8.5126*
18	11	-1.6202	20	11	-6.0038*
18	12	-13.0155*	20	12	-15.9860*
18	13	-0.8639	20	13	-5.2843*
18	14	9.7705*	20	14	5.7154*
18	15	8.0786*	20	15	3.8471*
18	16	4.7412*	20	16	0.2984
18	17	0.1269	20	17	-4.3299*
18	19	12.1056*	20	18	-4.4529*
18	20	4.4529*	20	19	8.3750*

*significance at or beyond the .05 level if the value of P(I,J) is greater than or equal to 3.3942.

Table 13

Rank Order and Significant Differences

Between Instructors on Dimension One



* = Significant difference between an instructor ranked above the * and all those ranked below the *, at or beyond the .05 level.

| = No significance between instructors.

Table 14

Rank Order and Significant Differences

Between Instructors on Dimension Two

Rank	Instructor	Instructor	Instructor	Instructor	Instructor
1	4				
	*				
2	1	1			
	|	*			
3	18	|			
	|				
4	11	11	11		
		|	*		
5		8	|		
		|	|		
6		16	16		
			|		
7			7		
			|		
8			13	13	
			|	*	
9			12	|	
			|	|	
10			17	17	
				|	
11				5	
				|	
12				20	
				|	
13				10	
				*	
14				14	
				*	
15				15	15
				|	*
16				6	
				|	
17				19	19
					*
18					9
					|
19					2
					*
20					3

* = Significant difference between an instructor ranked above the * and all those ranked below the *, at or beyond the .05 level.

| = No significance between instructors.

Table 15

Rank Order and Significant Differences

Between Instructors on Dimension Three

Rank	Instructor	Instructor	Instructor	Instructor
1	16	16		
2	17			
3	4			
4	12			
5	11			
6	13	13		
7		8	8	
8		18		
9		7		
10		1		
11		15	15	
12			14	
13			10	
14			19	
15			20	
16			5	5
17			9	
18			2	
19			6	6
20				3

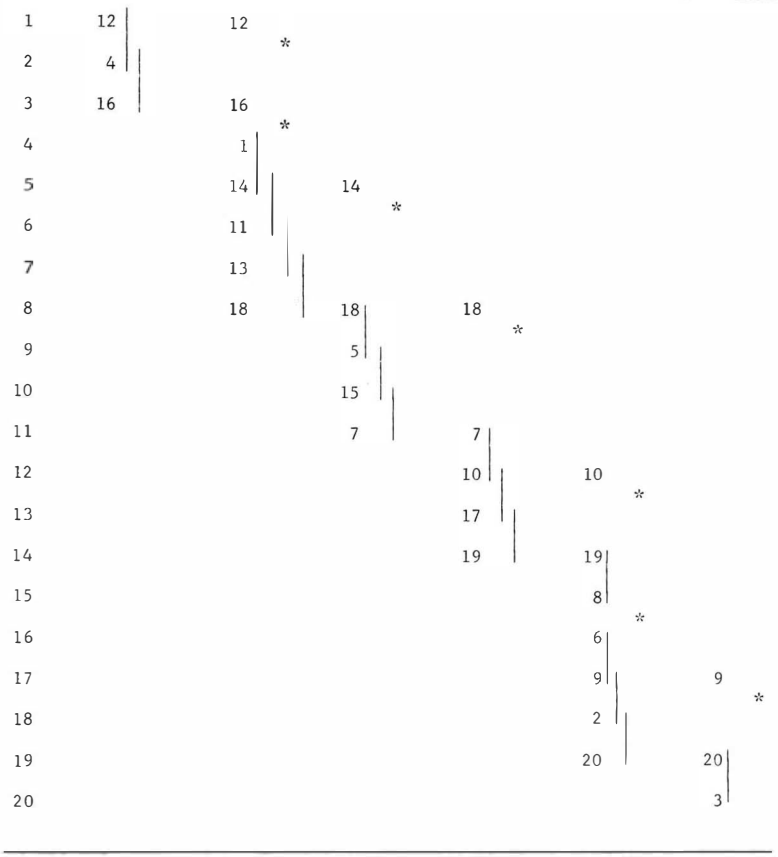
* = Significant difference between an instructor ranked above the * and all those ranked below the *, at or beyond the .05 level.

| = No significance between instructors.

Table 16

Rank Order and Significant Differences

Between Instructors on Dimension Four



* = Significant difference between an instructor ranked above the * and all those ranked below the *, at or beyond the .05 level.

| = No significance between instructors.

Table 17

Rank Order and Significant Differences

Between Instructors on Dimension Five

Rank	Instructor	Instructor	Instructor	Instructor
1	4	4		
		*		
2	16			
3	12			
4	17	17		
5		7		
6		1		
7		11		
8		13		
		*		
9		18	18	
			*	
10		8		
11		14		
12		19		
13		3	3	3
				*
14			15	
15			10	10
16				5
17				2
				*
18				20
19				6
20				9

* = Significant difference between an instructor ranked above the * and all those ranked below the *, at or beyond the .05 level.

| = No significance between instructors.

ranked number 3, etc. on dimension one. It can also be noted that instructors 12 and 4 are not rated significantly different but that instructor 4 is significantly different at or beyond the .05 level from instructor 5, and instructor 5 is significantly different from instructor 1. Therefore, instructors 4 and 12 are significantly different from all instructors below instructor 4, etc. These differences and standard deviations were derived from the table of pairwise preferences (see Table 12). It should be noted in this context that likelihood or probable ranks can be computed even though two instructors may never have been actually compared. For example, in Table 13, it may be seen that instructor 17 is ranked significantly above instructor 19, yet an inspection of Table 1 shows that 17 and 19 were never actually compared. But 17 and 19 shared in being compared to other instructors, and in one instance, both were compared to instructor 7 with 17 outranking 7 who, in turn, outranked 19. Such comparisons, along with the overall won-lost record of each instructor, allow a statement of significant differences to be made between 17 and 19.

In a simple rank order of instructors, without indicating significant differences (Table 18), it can be seen only that on dimension one, instructor 12 is ranked higher than instructor 4 and instructor 4 is ranked higher than instructor 5, etc.; and a similar expression of ranks is made in all other dimensions. Bradley (1952) states that rater agreement can be measured in a meaningful way only from one

instructor to another or from one dimension to another, and not from all instructors or dimensions combined. The comparisons of simple ranks, however, does provide for a fairly accurate assessment of instructor standing and gives an approximation of differences which can be useful when applied across dimensions for individual consistency or discrepancies between dimensions.

Evidence was obtained from these tables which clearly shows which instructors are consistently ranked by the students in the bottom or the top quartile of the faculty and which instructors fluctuate in rank among the various dimensions.

In the educational community, use can be made of this technique of instructor assessment. From the tables, a profile of the characteristics of an instructor, relevant to his teaching, can be drawn. For example (see Tables 13-18), instructor 4 is rated in the top quarter of the faculty on every dimension. This instructor can be described as one whom the students perceive as having a highly adequate knowledge of his subject matter, with his mastery on this dimension ranking significantly above all other instructors in the department; except for instructor 12 who received a higher, but not quite significantly different rank. In the presentation of course material, instructor 4 exceeds all other instructors in both rank and significance. Students view him as most effective in delivering lectures in both an interesting and intellectually stimulating manner, and as possessing the ability to relate important material to them.

Table 18

Simple Table of Rank Order of Instructors

On all Five Dimensions

Without Indicating Significant Differences

Rank	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5
1	12	4	16	12	4
2	4	1	17	4	16
3	5	18	4	16	12
4	1	11	12	1	17
5	10	8	11	14	7
6	11	16	13	11	1
7	13	7	8	13	11
8	18	13	18	18	13
9	17	12	7	5	18
10	8	17	1	15	8
11	7	5	15	7	14
12	20	20	14	10	19
13	16	10	10	17	3
14	15	14	19	19	15
15	14	15	20	8	10
16	19	6	5	6	5
17	6	19	9	9	2
18	9	9	2	2	20
19	2	2	6	20	6
20	3	3	3	3	9

Instructor 4's rating on adequacy in knowledge and presentation indicates that he is seen as being somewhat less able to maintain an adequate work load and to administer comprehensive, well-defined tests. The individual deficit in work load and testing can be clearly seen by a profile analysis of instructor 4 on dimension three where he ranked his lowest (third). A broader inspection reveals that there is no significant difference between the first five instructors on dimension three and only six significant differences in rank among all twenty instructors.

Apparently the students experienced some difficulty in differentiating the instructors' adequacy in knowledge and presentation. This probably reflects the instructors' difficulty in presenting the material clearly and then devising fair tests on what has been presented. It would appear a worthwhile venture to examine testing procedures among instructors. Perhaps students are reacting to inconsistencies among instructors in testing procedures or possibly to an extensive use of test items from the instructor manual.

Returning to the sample analysis of instructor 4, we note that he is consistently rated in the first quartile and his two strongest characteristics are presentation and approachability. Students perceive him as being very respectful of them, and they find communication with him easy outside of the classroom where they may seek his advice and discuss course work. Further, he is rated more than adequate in outlining course goals and in preparation for his lectures. The overall rating of instructor 4 is significantly above that of the majority of the instructors on the Psychology faculty.

It can be determined that some instructors consistently rank in the last quartile of the faculty population. Instructors 9, 2, and 6 are examples of this phenomenon. Three points are evident; (1) the consistency of rankings among the lowest ranked instructors apparently contributed strongly to the high rater agreement (.95), (2) it is easier to agree on the low ranked instructors than on the high ranked instructors, and (3) the students' perception of the low ranked instructors is that they are relatively bad instructors with less ability in all dimensions of instruction. Another possibility is suggested -- that there are consistently more bad instructors than there are good ones -- at least in the eyes of the beholder. A profile analysis of instructor 2, as an example of the low ranked instructor, reveals constant characteristics such as: (1) he does not have adequate knowledge of his subject matter, (2) his material is dated and not in the current trend, however, he is appraised significantly higher than instructor 3 on the same dimension, and (3) his presentation of the subject matter is perceived to be dull and not intellectually stimulating, with inadequate ability to transmit the course information.

Instructor 2 is relatively stronger (but still significantly lower than the average instructor in this study) in the appropriateness of class work load and testing procedures. Additionally, he is ranked higher than two of the faculty (but lower than 17) on preparation of lectures and defining objectives. Finally, his highest strength seems to be his interaction with the students. Three instructors are ranked lower (16 higher) in that area of

teaching which deals with fair and impartial treatment of students while maintaining respect and sensitivity for students as individuals. It would appear that any of the instructors consistently ranked in the bottom quartile are placed in a position where their teaching ability is being severely questioned by the consumer of their product.

Another profile which may be drawn from this study is one dealing with the erratic instructor. Such an instructor is defined as one who ranks high and low, as well as average, in the population on the various dimensions. An example of such an instructor is number 5. He ranks third in the population on his knowledge of the subject matter. Students believe that his material is kept relatively current and that he seems knowledgeable about his subject. He is significantly different on this dimension from all 17 professors ranked below him (only two are ranked above him). The students consider this his only strength and further agree that although his knowledge is sufficient and up-to-date, he is not communicating this knowledge to them. They do not evaluate his lectures as stimulating, rather they regard them as dull and uninteresting. The work load for the studnets is considered inappropriate and rigid. His tests are not well-defined and the students discriminate this instructor only as significantly different from the two instructors in the bottom of the population. The students rank instructor 5 somewhat higher on his ability to define course goals, but while this is one of his stronger attributes, he ranks barely above the fiftieth percentile in the population and is

significantly different from only three of the instructors below him. This instructor's greatest weakness is that he seems unapproachable both in and out of the classroom. He is significantly higher than only three other instructors on this attribute.

An instructional evaluation of faculty such as this one, where ranking and paired comparisons present a generally objective and precise picture of a given instructor, presents an overall view of the instructors' instructional prowess. Feedback from this evaluation should alert the instructor (and the administration) to the relative teaching ability of the faculty as viewed by the student. Since profiles can be drawn for each instructor and significant differences noted among them on each of the five dimensions, this type of evaluative process should become extraordinarily valuable as information for the instructor and could provide, with further systematic study of differences (in method, approach, attitude, etc.) between higher and lower ranked instructors, a greater understanding of the instructor's impact on the instructional process. Further, the objectivity and sharply differentiating ability of the instrument makes it useful as a part of the discriminative reward system for administrative purposes, both as a baseline measure and for assessment of change in the instructor along any basic dimension of instruction.

CHAPTER IV

SUMMARY

In the present study, a rating of professors as a variable in the learning process through a dimension rank order technique was made by 316 undergraduate students. An instructor evaluation instrument was used. This instrument was administered to students who were asked to rank order their professors on five basic dimensions of instruction.

The rank ordering demonstrated meaning in standard scores among instructors by a paired comparison transformation, and since parameters (a complete department) were used, absolute values were obtained. There was clear evidence to indicate that students could reliably and significantly discriminate between instructors on five basic dimensions of instruction. It was shown that students could significantly differentiate between instructors, revealing consistent preferences, i.e., rater agreement on the rank order of an instructor on a given dimension.

Profiles of instructors were derived from the constructed tables, thus it was determined that instructors could be assessed by this technique providing feedback to the instructor of his relative strengths and presenting objective data which could be used as a vital part of evaluating instructors on their teaching ability and in assessing the overall functioning of the instructional process in the University.

APPENDIX 1

RANK THESE FACTORS OF INSTRUCTIONAL EVALUATION ACCORDING TO IMPORTANCE, WITH RANK OF 1 BEING THE HIGHEST:

Skill - Ability - Presentation

_____ Essentially, what is meant here is the method or procedure by which an instructor gets important material across to the student in an interesting way, where clarity is maintained and the student is intellectually stimulated. This factor cuts across all other factors.

Stability Factor

_____ Does the instructor maintain a fair, reasonable, and appropriate work load in his assignments fairly evenly spread throughout the semester?

Organization

_____ Organization means things like lectures prepared, outlines presented, schedules, etc. This is sometimes referred to as course structure, preparation, etc.

Positive Response - Feedback

_____ Instructor compliments the student in class on good work; returns graded tests promptly with appropriate corrections; shows interest in student questions and encourages expression.

Instructional Image

_____ Instructor is characterized by an enthusiastic, friendly, flexible, and constructive approach.

---

1. WHAT DO YOU, AS A COLLEGE STUDENT AT V.C.U., BELIEVE ARE THE PREREQUISITES FOR GOOD COLLEGE INSTRUCTION? WHAT DO YOU THINK MAKES AN INSTRUCTOR MOST EFFECTIVE?

APPENDIX 2

INSTRUCTOR:_____ CLASS:_____

Please put a check by your class standing

Freshman _____

Sophomore_____

Junior_____

Senior_____

As part of a general study of instructional evaluation at VCU, you are asked to rank order the faculty of the psychology department for those individuals with whom you have taken courses here. First, please cross out the names of any that you have not had for a course. Then, rank by writing the instructor's name in the blanks starting with rank "1" as the best instructor on the faculty and then to to rank "21", the poorest instructor on the faculty. Then, alternate going from top to bottom -- the second best instructor having rank of "2", and the second worst instructor having the rank of "20", etc. Do not put your name or identification number on this sheet, but try to do the ratings honestly and carefully as this information will be invaluable.

1. The instructor has adequate knowledge of his subject matter.

2. Presentation: The instructor gets important course material across to the student in an interesting and intellectually stimulating manner.

3. The instructor maintains a reasonable and appropriate work load and administers comprehensive, well defined tests.

4. The instructor has well defined course goals, adequate preparation for the class and correlation between lecture material and the text.

5. The instructor is respectful of students and approachable by the students outside of the class time, i.e., they can come to see him.

Each of the above five dimensions were given on a separate page.

Names of faculty were listed in random order on separate pages following each individual listing of the five dimensions.

APPENDIX 3

TYPE II:

Each data card should contain in order the following information: EVALUATION, INSTRUCTOR, STUDENT, DIMENSIONS. EVALUATION, INSTRUCTOR, AND STUDENT numbers must be positive integers with ranges as indicated:

$$1 \leq \text{EVALUATION} \leq 999$$

$$1 \leq \text{INSTRUCTOR} \leq T$$

$$1 \leq \text{STUDENT} \leq 998$$

Score may be positive or negative, integer or real, decimal punched or not. The program assumes however that in any comparison the higher scores indicates the preferred treatment. Ties are ignored. Negative numbers must be indicated by a minus sign immediately preceding the leading non-zero digit.

The following is an example of TYPE II input:

EVALUATION	INSTRUCTOR	STUDENT	DIM. 1	DIM. 2	DIM. 3	DIM. 4	DIM. 5
1	3	1	1	10	9	5	8

Notice that more than one characteristic may be recorded on the input cards with this type data. Notice also that no instructor may occur more than once for a given evaluation and student.

ORDER OF DATA:

The input data must be sorted, as indicated below:

TYPE II, data must be sorted on evaluation, then student.

REFERENCES

Baker, P. & Butler, J. R. Personal communications, 1960.

Baker, P. Personal communication -- telephone conversation, March, 1970.

Bradley, R. A. Incomplete block rank analysis: on the appropriateness of the model for a method of paired comparisons. Biometrics, 1954a, 10, 375-390.

Bradley, R. A. Rank analysis of incomplete block designs II. Additional tables for the method of paired comparisons. Biometrika, 1954b, 41, 502-508.

Bradley, R. A. Rank analysis of incomplete block designs III. Some large-sample results on estimation and power for a method of paired comparisons. Biometrika, 1955, 42, 450-470.

Bradley, R. A. & Terry, M. E. Rank analysis of incomplete block designs I. The method of paired comparisons. Biometrika, 1952, 39, 324-335.

Brogan, H. O. Evaluation of teaching. Improving College and University Teaching, 1968, XVI (3), 191-192.

Coffman, W. E. Determining students' concepts of effective teaching from their ratings of instructors. Journal of Educational Psychology, 1954, 45, 277-386.

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

Dykstra, O., Jr. A note on the rank analysis of incomplete block designs -- applications beyond the scope of existing tables. Biometrics, 1956, 12, 301-306.

Dykstra, O., Jr. Rank analysis of incomplete block designs: A method of paired comparisons employing unequal repetitions on pairs. Biometrics, 1960, 16, 176-188.

Dykstra, O., Jr. Paired comparisons rank analysis of unbalanced data (Dykstra analysis). Unpublished Computer Program, 1970.

Eble, K. E. Project to improve college teaching. Academe, 1970, 4, 3-6.

Falk, G. The student views his professor. Improving College and University Teaching, 1968, XVI (3), 193-196.

Guilford, J. P. The method of rank order. Psychometric Methods. New York: McGraw-Hill, 1936, 244-262.

Hoffman, R. W. Students portray the excellent teacher. Improving College and University Teaching, 1963, XI (1), 21-24.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E. Correlation of teacher personality ratings and student ratings. Journal of Educational Psychology, 1963, 33, 415-441.

Isaacson, R. L., McKeachie, W. J., Milholland, J. E., et al. Dimensions of student evaluations of teaching. Journal of Educational Psychology, 1964, 55 (6), 344-351.

Kent, L. Student evaluation of teaching, In C. B. L. Lee (Ed.), Improving college teaching. Washington: American Council on Education, 1967.

Langen, T. D. F. Student assessment of teaching effectiveness. Improving College and University Teaching, 1966, XIV (1), 22-25.

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

McKeachie, W. J. Student rating of faculty. AAUP Bulletin, 1969a, 55 (4), 439-444.

McKeachie, W. J. Teaching tips--a guidebook for the beginning college teacher. Lexington, Massachusetts: D. S. Heath, 1969b.

Morton, R. K. What do students expect of a teacher? Improving College and University Teaching, 1963, II (1), 19-20.

Renner, R. R. A successful rating scale. Improving College and University Teaching, 1967, XV (1), 12-14.

Remmers, H. H. Rating methods in research on teaching. In N. L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand-McNally, 1963, 329-382.

Slobin, D. Y. & Nichols, D. G. Student rating of teaching. Improving College and University Teaching, 1969, XVII (4), 244-248.

Werdell, P. Course and teacher evaluation (2nd ed.) Washington: United States National Student Association, 1967.

VITA

Martha Louise Green was born October 2, 1946, in Lumberton, North Carolina. At the age of two, she moved to Dunn, North Carolina, now her permanent address. She attended public schools in Dunn and was graduated from Dunn High School in 1964. At the University of North Carolina at Greensboro, she majored in Psychology and received the degree of Bachelor of Arts in 1968.

During the summer immediately following graduation from college, she worked in the program evaluation section of the Veteran's Administration Hospital in Salisbury, North Carolina, under the direction of Dr. Mervyn K. Wagner. In the fall of 1968, she enrolled in the Clinical Psychology graduate program at Virginia Commonwealth University.

As a first year graduate student, she worked as a departmental assistant. During her second year, she served as research assistant to Dr. Joel R. Butler.

In April, 1970, she presented a paper at the Southeastern Psychological Association in Louisville, Kentucky, entitled, "Life History Antecedents of Delinquents and Murderers."

At present she is working in the Instructional Evaluation Program at Virginia Commonwealth University.

Permanent Address: 805 West Carr Street
Dunn, North Carolina 28334