May 1983

An Analysis of Student Evaluation of Selected Post Secondary Faculty Based on Student Perceptions of the Utilization of the Evaluation

Ann N. James
East Tennessee State University

Follow this and additional works at: https://dc.etsu.edu/etd
Part of the Education Commons

Recommended Citation

This Dissertation - Open Access is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
James, Ann Neblett

AN ANALYSIS OF STUDENT EVALUATION OF SELECTED POST SECONDARY FACULTY BASED ON STUDENT PERCEPTIONS OF THE UTILIZATION OF THE EVALUATION

East Tennessee State University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Ed.D. 1983
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark.

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Other

University Microfilms International
AN ANALYSIS OF STUDENT EVALUATION OF SELECTED POST SECONDARY FACULTY BASED ON STUDENT PERCEPTIONS OF THE UTILIZATION OF THE EVALUATION

A Dissertation
Presented to
the Faculty of the Department of Supervision and Administration
East Tennessee State University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
Ann Neblett James
May 1983
APPROVAL

This is to certify that the Graduate Committee of

ANN NEBLETT JAMES

met on the

___ 6th ___ day of ___ April ___ , 1983.

The committee read and examined her dissertation, supervised her defense of it in an oral examination, and decided to recommend that her study be submitted to the Graduate Council and the Dean of the School of Graduate Studies in partial fulfillment of the requirements for the degree Doctor of Education.

[Signatures]

Chairman, Advanced Graduate Committee

Signed on behalf of the Graduate Council

[Signature]

Dean, School of Graduate Studies
Abstract

AN ANALYSIS OF STUDENT EVALUATION OF SELECTED POST SECONDARY FACULTY BASED ON STUDENT PERCEPTIONS OF THE UTILIZATION OF THE EVALUATION

by

Ann Neblett James

The purpose of this study was to determine whether students' evaluation of faculty was affected by their perception of the utilization of the evaluation results. Three other variables were also investigated: faculty status, subject area of the class, and sex of the instructor.

Data for this study were collected from twenty classes taught by ten instructors at a private post-secondary institution in East Tennessee. Both instructors and their classes were randomly selected. The total number of students surveyed was 303. Two classes for each of the ten instructors were administered the evaluation instrument by the investigator. The control group was given oral instructions and information to the effect that the results of the evaluation survey would be used for the improvement of instruction. The experimental group was given the same oral instructions and information with additional treatment that the results would have input for personnel decisions.

The Student Instructional Report (SIR) was used as the instrument for collection of the student data. Data from the thirty-nine items on the SIR instrument were grouped by class into six factor categories: Factor 1, Teacher-Student Relationship; Factor 2, Course Objectives and Goals; Factor 3, Lectures; Factor 4, Reading Assignments; Factor 5, Course Workload; and Factor 6, Examinations.

Means for the thirty-nine items for each of the twenty classes were computed using the Statistical Package for the Social Sciences (SPSS). Grouped means in the six factor categories were used to perform t-tests for (1) control and
experimental classes by factor one to factor six, (2) classes taught by male and female instructors by Factor 1 to Factor 6, (3) classes taught by full-time and part-time faculty by Factor 1 to Factor 6, and (4) business or vocational classes and general studies classes by Factor 1 to Factor 6. Classes were used as the unit of analysis for the study.

Differences in student ratings due to different instructions on the intended uses of the results were not statistically significant at the .05 level of significance. Data were analyzed for experimental and control classes for each of the factor categories as well as for global rating items 38 and 39. No consistent pattern emerged in the data analysis. Although not at a statistically significant level, negative differences indicating the control classes had a higher summarized mean occurred for five factors in a summary of the mean scores of the six factor categories. Five teachers received higher mean scores for the control class for item 38 dealing with overall value of the class to the student. Six teachers received higher mean scores for the control class for item 39 dealing with overall teacher effectiveness.

Summarized data from all classes taught by part-time faculty showed higher mean scores in four of the factor categories. Factor 1, teacher-student relationship, showed the highest mean for part-time faculty and the greatest difference in the t comparisons with full-time faculty. None of the differences from the comparison of full-time and part-time faculty were significant at the .05 level in any of the six factor categories.

When summarized mean scores for business or vocational classes and general studies classes were compared, t values did not indicate differences significant at the .05 level. General studies classes did receive higher mean scores for all the factor categories except Factor 4 dealing with reading assignments.

Male teachers received higher mean scores for all the factor categories based on summarized mean scores for all their classes. When compared to the summarized mean scores for classes taught by female instructors, the differences were not significant at the .05 level.

The data from the study seem to indicate that the students at the institution surveyed were discriminating and relatively free from the influences of the variables in this study in their evaluative ratings of faculty and instruction.
INSTITUTIONAL REVIEW BOARD

This is to certify that the following study has been filed and approved by the Institutional Review Board of East Tennessee State University.

Title of Grant or Project AN ANALYSIS OF STUDENT EVALUATION OF SELECTED POST SECONDARY FACULTY BASED ON STUDENT PERCEPTIONS OF THE UTILIZATION OF THE EVALUATION

Principal Investigator Ann Neblett James

Department Supervision and Administration

Date Submitted January 28, 1983

Institutional Review Board Approval Chairman [Signature]
DEDICATION

To my children,

Betsy, Trigg, Jimmy, Becky, and Sarah
who have always supported me in my belief that education
has prime value for our lives.

To my grandchildren,

Ashley, Emily, Rachel, Mindy, and Audrey
who are just beginning their journey in our footsteps.
ACKNOWLEDGMENTS

The writer wishes to express heartfelt thanks to Dr. William Pafford, who chaired her committee and directed her dissertation efforts. Without his guidance, support, encouragement, and unselfish giving of his valuable time, this dissertation would have been impossible.

Grateful appreciation is also extended to Dr. Charles Burkett, Mr. Herbert Armentrout, Dr. Floyd Edwards, and Dr. William Fowler for serving on her graduate committee and for offering so freely their assistance and support.

Sincere thanks are also expressed to Madaline Jenkins who typed this manuscript so willingly, and David Baker who assisted tirelessly in the computer analysis of the data.

Finally to my sons-in-law Frank, Todd, and Randy thanks for being so tolerant of my eccentricities while writing of this dissertation was in progress.
CONTENTS

APPROVAL .............................................. ii
ABSTRACT ................................................. iii
INSTITUTIONAL REVIEW BOARD ........................................... v
DEDICATION ............................................ vi
ACKNOWLEDGMENTS ........................................ vii
LIST OF TABLES ........................................ xi

Chapter

1. INTRODUCTION ............................................ 1
   The Problem ........................................... 4
   Statement of the Problem ................................ 4
   Subproblems ............................................ 4
   Limitations .............................................. 4
   Hypotheses .............................................. 5
   Definitions of Terms .................................. 6
   Business or Vocational Subjects ...................... 6
   Evaluation .............................................. 6
   Evaluation Instrument ................................... 7
   Evaluation Report ....................................... 7
   Evaluators .............................................. 7
   Faculty Status ......................................... 7
   Formative Evaluation Procedures ...................... 7
   General Studies Subjects ................................ 7

viii
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Instructional Report</td>
<td>8</td>
</tr>
<tr>
<td>Student Perception</td>
<td>8</td>
</tr>
<tr>
<td>Summative Evaluation Procedures</td>
<td>8</td>
</tr>
<tr>
<td>Utilization</td>
<td>8</td>
</tr>
<tr>
<td>Organization of Study</td>
<td>8</td>
</tr>
<tr>
<td>2. REVIEW OF THE LITERATURE</td>
<td>10</td>
</tr>
<tr>
<td>Evaluation Utilization</td>
<td>10</td>
</tr>
<tr>
<td>Sources of Information in Evaluating Teaching Effectiveness</td>
<td>14</td>
</tr>
<tr>
<td>Variables That Affect Student Evaluation</td>
<td>21</td>
</tr>
<tr>
<td>3. RESEARCH PROCEDURES</td>
<td>30</td>
</tr>
<tr>
<td>Description of the Sample</td>
<td>30</td>
</tr>
<tr>
<td>Procedures</td>
<td>31</td>
</tr>
<tr>
<td>Description of the Instrument</td>
<td>34</td>
</tr>
<tr>
<td>4. ANALYSIS OF THE DATA</td>
<td>37</td>
</tr>
<tr>
<td>Testing of the Hypotheses</td>
<td>38</td>
</tr>
<tr>
<td>Comparison of the Control and Experimental Groups</td>
<td>40</td>
</tr>
<tr>
<td>Comparison of the Classes Taught by Full-time Faculty and Classes Taught by Part-time Faculty</td>
<td>47</td>
</tr>
<tr>
<td>Comparison of Business or Vocational Classes and General Studies Classes</td>
<td>49</td>
</tr>
<tr>
<td>Comparison of the Classes Taught by Male Teachers and the Classes Taught by Female Teachers</td>
<td>49</td>
</tr>
<tr>
<td>Discussion of Results</td>
<td>52</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5. SUMMARY, CONCLUSIONS AND IMPLICATIONS</td>
<td>55</td>
</tr>
<tr>
<td>Summary</td>
<td>55</td>
</tr>
<tr>
<td>Conclusions</td>
<td>58</td>
</tr>
<tr>
<td>Implications</td>
<td>59</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>50</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>70</td>
</tr>
<tr>
<td>A. OVERVIEW REMARKS OF MEETING WITH BRISTOL COLLEGE AND HANDOUT TO FACULTY</td>
<td>71</td>
</tr>
<tr>
<td>B. EXPLANATIONS TO CONTROL AND EXPERIMENTAL GROUPS</td>
<td>74</td>
</tr>
<tr>
<td>C. PERMISSION LETTERS FROM EDUCATIONAL TESTING SERVICE AND BRISTOL COLLEGE</td>
<td>76</td>
</tr>
<tr>
<td>VITA</td>
<td>79</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>SUMMARY OF TEACHER MEANS FOR CONTROL (A) AND EXPERIMENTAL (B) GROUP BY FACTOR</td>
<td>39</td>
</tr>
<tr>
<td>2.</td>
<td>DIFFERENCES IN THE MEANS FOR EXPERIMENTAL AND CONTROL CLASSES FOR EACH INSTRUCTOR</td>
<td>41</td>
</tr>
<tr>
<td>3.</td>
<td>t SCORES FOR CONTROL AND EXPERIMENTAL CLASSES BY FACTOR</td>
<td>43</td>
</tr>
<tr>
<td>4.</td>
<td>TEACHER MEANS FOR SIR ITEM 38 BY CLASS</td>
<td>44</td>
</tr>
<tr>
<td>5.</td>
<td>TEACHER MEANS FOR SIR ITEM 39 BY CLASS</td>
<td>46</td>
</tr>
<tr>
<td>6.</td>
<td>t SCORES FOR CLASSES TAUGHT BY FULL-TIME AND PART-TIME FACULTY BY FACTOR</td>
<td>48</td>
</tr>
<tr>
<td>7.</td>
<td>t SCORES FOR BUSINESS OR VOCATIONAL AND GENERAL STUDIES CLASSES BY FACTOR</td>
<td>50</td>
</tr>
<tr>
<td>8.</td>
<td>t SCORES FOR CLASSES TAUGHT BY MALE AND FEMALE TEACHERS BY FACTOR</td>
<td>51</td>
</tr>
</tbody>
</table>
CHAPTER ONE

Introduction

Evaluation is a multifaceted concept in the minds of the cultural milieu of modern day society. It may be conceived as a process or a product, as a formal, planned activity using a written instrument with data results, or as an informal decision made in private based on almost anything.

Teaching and learning are not solitary activities. Evaluations and judgments concerning the effectiveness of the process, the product, the setting, and the people involved are inherent in education as it has been formalized and has evolved from earliest times to its present day priority in the lives of the majority of the people in this country. "In higher education we have always evaluated teaching, and if anything we are likely to do even more of this in the future" (Cashin, 1978:3).

To evaluate or not to evaluate, then, does not seem to be a relevant question in the present educational arena of post-secondary institutions. A more realistic question and one that will be addressed in this study concerns the utilization of formal evaluation results and whether a prior knowledge of what the utilization will be affects the response of the student evaluators.

1
In an early report on evaluation utilization Weiss (1972:318) stated that if evaluation results were not used to make program decisions then evaluation had "failed in its major purpose." Over a dozen reasonably distinct purposes for utilization of teacher evaluation have been suggested, such as improving teacher performance, aiding administrative decisions, guiding students in course selection, meeting state and institutional mandates, promoting research on teaching, and "the like" (Millman, 1981).

Two major broad utilization categories seemed to emerge from the abundance of information on the subject. One was utilization in a formative mode—i.e., helping faculty improve their performance by providing data, judgments, and suggestions that have implications for what to teach and how to do so. The second major utilization mode was a summative one—i.e., serving administrative decision making with respect to hiring and firing, promotion and tenure, assignments and salary.

Who should evaluate and how, since the utilization of the results could and should have such far reaching implications, are questions that continue to plague those responsible for such decisions. Socrates talking in the market place or Aristotle walking in the Lyceum gardens gathered about them whoever was attracted to their teaching. The simplicity of their effectiveness, to be able to attract students, is antiquated in the complexity of organized
education. Many groups today are concerned about faculty evaluation utilization and should have input for some utilizations. They may include faculty peers, administrators, students, parents and the public, professional evaluators, and the teacher himself.

One of the most frequently used groups for faculty evaluations is the student. It has been reported in various studies that student ratings are both valid, reliable, and useful, and unreliable, invalid, and useless. In fact, however, student ratings tend to be the only tangible source of instructional evaluation information in the majority of colleges and universities, both here and abroad (Thorne, et al., 1976)

Particular stress has been placed on those responsible for faculty evaluation to choose the right instrument to give the most penetrating analysis. The evaluator has many tools available. Types of student ratings range from formally developed, sophisticated questionnaires printed on optically scanned answer sheets to informal forms made by an instructor for his class. Questions require answers that range from an objective check to an extensive discussion or some combination of the two.

Currently the evaluation of faculty is an extremely important activity. It affects many aspects of the educational process and the entire career of many individuals. Evaluation utilization, then, is not a trivial
matter. Since student evaluations of faculty are so widespread, their perceptions of evaluation utilization would appear to be a subject worthy of attention.

The Problem

Statement of the Problem

The problem of the study was to determine whether students' evaluation of faculty was affected by their perception of the utilization of the evaluation results.

Subproblems

The following subproblems have been developed:

1. To determine whether students' evaluation of faculty was affected by whether the faculty member was a full-time instructor or a part-time instructor.

2. To determine whether students' evaluation of faculty was affected by the subject area of the class.

3. To determine whether students' evaluation of faculty was affected by the sex of the instructor.

Limitations

The following were considered to be limitations of the study:

1. This study was limited to a random sample of students and faculty from one post-secondary institution in Upper East Tennessee.

2. The study involved two classes taught by each of ten instructors at the institution researched. No attempt
was made to generalize to other classes taught by these ten instructors or to other classes or instructors at the institution involved in the study.

3. The study was limited to the scope of the Student Instructional Report (SIR) instrument developed by Educational Testing Service of Princeton, New Jersey.

4. The study was limited to post-secondary students. No attempt was made to generalize to other educational levels.

5. The research was limited to data collected at the end of Winter Quarter 1983, prior to examination time.

Hypotheses

Hypothesis I

Student evaluations of faculty using the mean scores from the SIR evaluation instrument will show significant statistical differences at the .05 level between students that perceive the evaluation results will be used for improvement of instruction and students that perceive the evaluation results will be used for personnel decisions.

Hypothesis II

Student evaluations of faculty using the mean scores from the SIR evaluation instrument will show significant statistical differences at the .05 level between classes taught by full-time faculty and classes taught by part-time faculty.
Hypothesis III

Student evaluations of faculty using the mean scores from the SIR evaluation instrument will show significant statistical differences at the .05 level between business or vocational classes and general studies classes.

Hypothesis IV

Student evaluations of faculty using the mean scores from the SIR evaluation instrument will show significant statistical differences at the .05 level between classes taught by male teachers and classes taught by female teachers.

Definitions of Terms

Business of Vocational Subjects

Those subject areas whose main objective is to prepare the student for a targeted job or business occupation are business or vocational subjects.

Evaluation

An evaluation is the process that involves (a) posing questions about the purpose, implementation and consequences of . . . programs and people and (b) systematically collecting and analyzing data concerning these questions, where both of these activities are intended to facilitate judgment about the worth of such programs (Weiner, et al., 1977:2).
Evaluation Instrument

A rating form, designed to elicit information by posing pertinent questions to be marked by the student from answers on some continuum from very low to very high, is an evaluation instrument.

Evaluation Report

Often referred to as an evaluation, an evaluation report is the product of an evaluation.

Evaluators

Professionals or nonprofessionals who participate in the evaluation process and/or make judgments from evaluative data are evaluators.

Faculty Status

Faculty status denotes whether an instructor is teaching full-time (at least 15 quarter hours) or part-time (3 to 12 quarter hours).

Formative Evaluation Procedures

Professional growth decisions based on evaluation reports are formative evaluation procedures.

General Studies Subjects

Those subject areas in which the main objective is to impart general knowledge about the subject are general studies subjects.
Student Instructional Report

The Educational Testing Service of Princeton, New Jersey, developed the student evaluation instrument known as the Student Instructional Report. The acronym for this instrument is SIR.

Student Perception

A mental image or concept of the student is his perception.

Summative Evaluation Procedures

Personnel decisions based on evaluation reports are summative evaluation procedures.

Utilization

Utilization refers to evaluation information considered as an influence in "making decisions, substantiating previous decisions or actions, or establishing or altering attitudes" about a variety of items (Alkin, et al., 1979: 232).

Organization of the Study

The study was organized into five chapters. Chapter 1 contains an introduction to the study, statement of the problem, subproblems, limitations of the study, and hypotheses. Definitions of terms and organization of the study are also included in Chapter 1.

A review of related literature is presented in Chapter 2. The readings deemed most significant were categorized
and reported in this chapter.

Chapter 3 describes the sample used in the study, the research procedures, and a description of the instrument used for the collection of the data.

An analysis of the data is presented in Chapter 4.

Chapter 5 contains the summary, conclusions, and implications of the study.
CHAPTER TWO

Review of the Literature

A search of the literature yielded information which was relevant to this study in the following areas: (1) utilizations of formal educational evaluations, (2) sources of information in evaluating teaching effectiveness, and (3) variables that affect student evaluations. The readings deemed most significant were categorized accordingly and are reported in this chapter.

**Evaluation Utilization**

The literature in the area of utilization of evaluation results revealed the fact that formal evaluation data were being used in various ways. In contrast to the popular feeling that evaluation results were underutilized, Patton et al. (1978) conducted a follow-up study of twenty health program evaluations. They found that results were in fact being used, but not in the general sense of having an "immediate and concrete effect on specific decisions and program activities." Instead, evaluation results provided the decision makers one additional piece of information, "thereby permitting some reduction in the uncertainty within which any decision maker inevitably operates" (Patton et al., 1978:143-145).
Alkin, Daillak, and White (1979) distinguished two views of evaluation utilization: a mainstream view that looked for the specific impact of an evaluation on subsequent decisions; and a broader, alternative view that examined the numerous direct and indirect effects an evaluation could have on an organization. The consensus that evaluation results were not being used may have been due to the bias of the mainstream perspective according to the study. Evaluation results may have already influenced programs, but in more subtle ways than the search for static factors affecting utilization would have suggested.

Based on interviews with 116 federal policy analysts, A. J. Meltsner (1976) conceptualized a typology of evaluators. He argued that the personal factor was a primary determinant of use—i.e., the presence of a person who cared about the evaluation and its results. This work was similar to Patton's (1978) and other theorists in philosophy.

In a review article summarizing a "five year series of studies which had systematically examined the relationship among the characteristics of an evaluator, an evaluation report, evaluation audience characteristics, and audience responses," Newman et al. (1980:20) reported some interesting factors that affected evaluation utilization. They asked subjects representing a variety of evaluation audiences to read and respond to simulated evaluation reports. Although it may be questioned whether or not generalizations
may be made from such studies, the results suggested several important points concerning the utilization of evaluation results. First, both the title and sex of the evaluator could affect audience reactions. Second, the use of jargon and data could affect audience ratings of technicality and difficulty. "Generally, reports containing both jargon and data were rated more useful. . . ."

Third, the "... audience's perceived need for evaluative information in a particular area" affected utilization (Newman et al., 1980:33).

Several recent doctoral dissertations have addressed the question of evaluation utilization. Carlson (1974) studied the relationship of utilization, as measured by a self-developed index, and three variables; the clarity of organizational goals and objectives; the number of individuals necessary for approval of a recommendation; and the status of the evaluator. Positive relationships existed between utilization and clarity and between utilization and the internal status of the evaluator.

To determine factors affecting the utilization of forty-seven completed Title IV-A evaluations, Dickey (1979) interviewed project directors, read final evaluation reports, and collected archival data. She concluded that the likely explanation of underutilization lay in factors related to the natural resistance to change and to the dissimilarity of the academic and real worlds, rather than to an
impoverished state of evaluation art.

In another doctoral dissertation study, Weeks (1979) sought to identify which of three variables correlated most highly with utilization. The variables studied were the organizational location of the evaluator; the decision-making context; and the methodological practices employed. The negative correlation between research design and utilization suggested that decision-makers had a slight preference for more qualitative forms of data analysis.

In a recent paper presented at the American Educational Research Association, Haenn (1982) reviewed the literature and discussed three sets of factors which inhibit information use. The first was organizational characteristics; the structure, climate, and politics of organizations may have limited the effective utilization of information. A second type of inhibiting factors was the personal characteristics of users, whose information needs, interests, and abilities may have affected the use of information. The third set of factors encompassed methodological characteristics of evaluations and reporting, including both the role of the evaluation and the characteristics of its reporting. Haenn summarized his review of the literature by noting that, "the literature is filled with reasons why information may not be effectively utilized."

The common, striking theme that emerged from a review of the literature in the area of evaluation utilization was
that the people involved were the key factors to utilization. This common thread ran through all the studies—the caring and/or status of the evaluator(s), the political tone of the organization, and the personal characteristics of the users and their perceived needs, how those being evaluated felt about the evaluation process, the subtle impact that evaluations had on decision makers, and how the evaluative data were presented. The technical state of the art of evaluating was well developed and sophisticated. The wealth of instruments available for this purpose was overwhelming. The negative correlation between research design and utilization was enlightening. The personal factors, the people involved, emerged as highly important.

Sources of Information in Evaluating Teaching Effectiveness

Colleges, universities, elementary and secondary schools were reported struggling with the dilemma of providing objective data for decisions about teachers and instruction. The then Vice-President of Academic Affairs of East Tennessee State University, Donald Goodman, pointed out the problem in a student publication when he said, "with the absence of good evaluations, our decisions (about instructors) become subjective" (Kirkland, 1982:1). A Miami union representing the Miami area's 23,500 teachers recently cooperated with the Date County School System to
put in place an intensive new evaluation plan that will provide objective data in the form of a prescribed checklist which principals must use in required observations of new nontenured and tenured teachers (Toch, 1982).

In spite of the wealth of evaluation instruments, there seemed to be no single indicator of educational effectiveness that could not be criticized for some theoretical or practical fault according to E. Grady Bogue (1982) writing on college administration in the 1980s. He further contended that it was difficult to assess effectiveness unless purpose had first been clarified. One of the problems, as he saw it, was that the technical and philosophic difficulties with educational evaluation tended to immobilize while educational institutions looked for the perfect process.

Evaluation specialists and faculty members consulted by writers in "The Chronicle of Higher Education" on the question of whether fair and useful evaluations of professors by students could be done had several common conclusions (Perry, 1982:19). Represented in the group were William E. Cashin, an educational-development specialist at the Center for Faculty Evaluation and Development at Kansas State University; Lawrence M. Aleamon, Director of the Office of Instructional Research and Development at the University of Arizona; and John A. Centra, program administrator at the Educational Testing Service. All, although
they represented different approaches to the evaluation process, agreed that a well constructed instrument, administered fairly and uniformly with results interpreted intelligently with follow-up, were prime considerations.

A number of well-tested student rating questionnaires are nationally available to colleges and universities. Three emerged as the most widely used. The Instructional Development and Effectiveness Assessment (IDEA) System developed and copyrighted by the Center for Faculty Evaluation and Development at Kansas State University has been used at more than 350 colleges and universities. It asks students to evaluate their instructor in terms of how much progress they believe they made toward the stated objectives of the course. The Student Instructional Report (SIR) developed and copyrighted by Educational Testing Service of Princeton, New Jersey has been used at over 400 colleges and universities. It asks for students' reactions to their instruction in a rating format using thirty-nine questions. The Arizona Course/Instructor Evaluation Questionnaire (C.I.E.Q.) developed and copyrighted by Lawrence M. Aleamoni at the University of Arizona has been used on about 125 campuses. It asks students to respond to twenty-one positive and negative statements by marking Agree Strongly, Agree, Disagree, or Disagree Strongly. The instructor has the option of selecting up to forty-four additional questions from a
catalog. Each of these evaluation instruments and their supporting institutions have continuing active research programs as well as national data bases.

Bearing out the need for standardization, whatever the method used, one of the conclusions reached by a resource panel providing information for Phi Delta Kappa's newsletter, Practical Applications of Research (1982:3) was that "although it is uncertain how much procedural differences affect the evaluation results, the need for some standardization seems apparent."

In spite of the presence of well developed and continuously improved instruments, each with their standardized procedure for administration of the evaluations and interpreting the data, faculty, students, and administrators persist in clamoring for a better evaluation. In a student editorial, a student wrote in "The East Tennessean," "Unfortunately, if the evaluations are conducted under the same conditions as in previous years, the ultimate result will be a mockery of what the evaluations were intended to do . . ." (McClellan, 1982:28). In a speech made before members of the Student Government Association, President Ronald Beller of East Tennessee State University said in October, 1982 that faculty evaluations were ten years behind the times, that faculty should be accountable for what occurred in the classroom, and that present evaluation practices did not insure this.
Articles have also appeared even in secular magazines demanding better teacher evaluations or calling for effective teacher evaluation programs (Kaercher, 1982). Declining public school enrollments, poor showings on standardized tests of basic skills, and reduced budgets have made teacher competency and effective evaluation a prime topic among parents, educators, and legislators.

Out of all this furor, some research conclusions emerged as pointed out by Robert Stake (1976) in a study done by him for the Organization for Economic Cooperation and Development. No one method of evaluating instruction could be suitable for all situations. The information needs varied. The audiences have had different expectations and standards. Evaluators had different styles, which were more or less useful to different clients.

The writing of Michael Scriven (1967) has been influential in identifying basic dimensions of evaluation. His paper identified six dimensions starting with a distinction between the goal of evaluation (to indicate "worth") and the roles of evaluation (the different reasons and circumstances for which we need to know the "worth"). The most useful distinction here may have been between the users of the evaluation findings. In this regard, when the cook tastes the soup it is formative evaluation and when the guest tastes the soup it is summative. The key is not as much when as why. What is the information for, for further preparation and
correction, or for savoring and consumption? Both lead to decision-making, but toward different decisions" (Stake, 1976:19).

The problem of evaluating faculty in a cost efficient manner was seen as a major problem at many colleges and universities. A system using multiple sources such as that advocated by the Higher Education Council of the National Education Association (Kronk & Shipka, 1980) may be ideal and produce the best results, yet be costly and difficult to implement. A model that would seem to counteract these criticisms was implemented in late 1978 at the Louisiana State University School of Dentistry. A paper presented at the Mid-South Educational Research Association in 1979 extolled the results of the research (Boozer, et al., 1979). A standardized rating form using a Likert scale was used to assess instructors numerically. The Curriculum Committee interviewed five randomly-selected students who reported on each instructor in a structured interview technique. An unstructured interview method was used by the Dean to communicate with five other randomly-selected students. The results of the three analyses were used to consult with the instructor.

In a study of practices in collecting information for evaluating classroom teaching performance in liberal arts colleges, Seldin (1975) reported the frequency of the use of fifteen sources of information.

1. Chairman evaluation (tied for first)
2. Dean evaluation (tied for first)
3. Committee evaluation
4. Colleagues' opinions
5. Systematic student ratings
6. Self-evaluation
7. Scholarly research and publication
8. Informal student opinions
9. Course syllabi and examinations
10. Classroom visits
11. Student examination performance
12. Enrollment in elective courses
13. Grade distribution
14. Long-term follow-up of students
15. Alumni opinions

Miller (1972) suggested using only four of these: classroom visits, course syllabi and examinations (which he called teaching materials and procedures), self-evaluation, and systematic student evaluation.

It should be clear from the foregoing review that no single source of information can be recommended without reservation as the basis for evaluating teaching. Neither is education postponing the task until a problem-free source of information is developed. "Teaching must be evaluated, and in fact has been evaluated for decades using sources far more 'disadvantaged' than those discussed" (Cashin, 1978:16). The question, then, is
not whether to evaluate but how and why with the why emerging as most important for choosing the process and analyzing the product.

**Variables That Affect Student Evaluation**

After a search of the literature, some variables which affect student evaluations have been researched to the extent that they are sufficiently well established to be controlled for in at least two evaluation systems, SIR and IDEA. Normative data for comparison purposes were available both from Educational Testing Service and the Center for Faculty Evaluation and Development at Kansas State University (Aubrecht, 1979; SIR Comparative Data Guide, 1979). National normative data were available for extraneous variables not under the control of the teacher in several categories. Among these were class size and initial student motivation. Smaller classes received higher ratings with correlations reported between student ratings and class size generally ranging from negative .10 to negative .30. Some studies have reported a nonlinear relationship with medium size classes receiving lower ratings than either smaller or larger classes (Feldman, 1978; Costin, et al., 1971). Correlations of .42 and .48 between students' initial liking for the subject and global and overall evaluation items in student ratings have been reported by Doyle and Whitely (1974).
Some situational conditions could also be considered extraneous variables. Students tended to rate instructors higher if the ratings were not anonymous or if the instructor was present while they were completing the forms (Aubrecht, 1979; Feldman, 1979).

These situational conditions could be controlled by having standardized procedures throughout the evaluating institution as pointed out earlier in this literature search (Perry, 1982).

Studies correlating grades and student ratings were numerous. Most of the studies correlating grades and student ratings reported correlations from the mid .10's to just below .30. There was some evidence that grades and ratings may have strong positive correlations in some classrooms, no apparent relationship in others, and even negative correlations in others (Feldman, 1976). Most studies, however, have found a small but significant relationship between student's grade (or expected grade) in a course and ratings of instruction (Spencer, 1965; University of South Florida, 1965; Centra, 1964). Elliott (1950) suggested that a positive correlation between ratings and grades would be expected if the instructor "teaches to" the better students in the class and negative correlation if the instructor "teaches to" the poorer students. Centra (1976b) did conclude that the moderate relationship between
students' expected grade and their ratings did not prove that students reward easy-grading teachers with higher ratings.

Teacher characteristics and their relationship to student ratings have been researched quite extensively. The relationship between student ratings and either academic rank and research productivity seemed to be positive but quite weak (Aubrecht, 1979; Aleamoni and Yeimer, 1973; Centra, 1976b; Centra, 1979). Instructors with high ratings seemed to differ from those with low ratings on measures of communication ability (Kulik and McKeachie, 1975; Costin and Grush, 1973). Costin and Grush also showed a significant correlation between student ratings and teacher energy level characterized by rapid work and accomplishing a great deal. There was added support for the variable of professor clarity and organization relating to teacher ratings in a recent study by Carlson and Bensinger (1981). When lecturers' presentations were highly expressive, student ratings were not sensitive to either differences in content or to students' actual test performances as reported in two studies done by Ware and Williams (Ware and Williams, 1975; Williams and Ware, 1977). They called this lack of sensitivity to substance of highly expressive presentations the "Dr. Fox effect."
Time does not seem to affect student ratings of instructors. In an early study of alumni ten years after graduation, respondents agreed with on-campus students in their average ratings of the same instructors (Drucker and Remmers, 1950). In that study, Drucker and Remmers reported correlations ranging from .40 to .68 on ten teacher traits. Centra (1973a) supported the previous study by reporting that judgments of teachers by their students at the end of a course were fairly permanent. The agreement between current students and alumni of five years was substantial, particularly in identifying teachers at the extremes.

Student variables such as students' curriculum, year in college, overall grade point average, student needs and a priori student theories have been studied. Students' curriculum or year in college had little or no relationship to their course ratings (Centra and Creech, 1976; Spencer, 1965). Neither did students with higher overall grade-point averages rate instructors differently than did lower-achieving students (Guthrie, 1954; Centra and Creech, 1976). Student needs were found to influence some items on the Purdue Rating Scale (Rezler, 1965); however, there were no significant relationships among students' ratings of the professor and their ratings of themselves as students in a recent Middle Tennessee State University study (Carlson and Bensinger, 1981). Whitely and Doyle
(1976) reported that students organize and evaluate their experience of an instructor with respect to a priori behavioral categories.

Ratings of students in different subject areas did vary. The differences in ratings among subject areas were quite significant in Centra and Creech's 1976 study. The natural sciences received the lowest ratings followed by the social sciences. Humanities received the highest ratings. Another SIR study (Centra, 1972) indicated that courses in the natural sciences compared to those in the humanities, social sciences, and education were perceived by students as having a faster pace, as being more difficult, and less likely to stimulate student interest. Feldman (1978) found that teachers in the humanities, fine arts and languages received higher ratings than teachers in social or physical sciences, mathematics and engineering. No studies were found on ratings in business and vocational subjects, although SIR has comparative data on Business and Management courses (SIR Comparative Data Guide, 1979).

Conflicting results have been obtained when relating the sex of the student and/or sex of the instructor to students' evaluations of instruction. Male and female students did not generally differ in their rating of instructors (Spencer, 1965; Aleamoni and Thomas, 1977; Centra, 1976b). In addition Costin et al. (1971) cited seven studies that reported no differences in overall
ratings of instructors made by male and female students or in ratings received by male and female instructors. Conversely, Walker (1969) found that female students rated female instructors significantly higher than they rated male instructors. In addition, Aleamoni and Hexner (1980) cited studies that reported female students rated instructors higher on some subscales of instructor evaluation forms than did male students. McKeachie and Lin (1971) reported that with male teachers high warmth was more important to the achievement of female students than of male students. Apparently sex of the student sometimes interacts with sex of the teacher to produce somewhat different rating patterns.

Conflicting results have also been found when comparing teaching experience to student ratings. In a very early study, it was found that teachers with less than five years of experience tended to be rated lower than teachers with more than eight years experience (Remmers, 1929). In another study, Heilman and Armentrout (1936) found no significant relationship, whereas Rayder (1968) reported a negative relationship. Graduate teaching assistants as a group were consistently rated lower than were full-time teachers (Centra, 1976). In a very comprehensive study done by Centra and Creech (1976), using the SIR instrument, teachers with between three to twelve years of experience received the highest mean rating: 3.83.
First year teachers clearly received the lowest rating mean; 3.54. The decline in teaching effectiveness after twelve years was significant according to student ratings.

Educational Testing Service provides normative data giving the percentile rank of item means for full-time and part-time faculty (SIR Comparative Data Guide, 1979). Studies have been done on credit-hour teaching load and student ratings. Teachers with teaching loads of thirteen or more hours were rated higher than any other group. The lowest ratings were for teachers with four to six hour loads (Centra, 1976b). Some of the teachers carrying the low teaching loads could fall into the part-time category but not necessarily so. Others in this low teaching load category could have been doing research or assuming administrative duties.

A critical question in the present study was whether students evaluate instruction differently depending on their perception of the intended use of the results. Wherry (1952) expected ratings to be more favorable when used for administrative purposes. According to Wherry, the more favorable ratings he reported were due to a bias in recall which occurred in those situations where raters tended to recall their more favorable perceptions of the ratee's performance. He argued that ratings obtained under conditions where resulting administrative action could affect the ratee would be less accurate than those
obtained for the improvement of instruction or methods. Taylor and Wherry (1951) reported more favorable ratings in a military setting when raters (U. S. Army Officers) were told that the results would be used for administrative purposes. Sharon and Bartlett (1969) compared student ratings in a college teaching situation and found that ratings were slightly more favorable when raters were informed that the results might be used administratively. The ratees in the Sharon and Bartlett study were fourteen graduate teaching assistants; therefore, the results may not be applicable to full-time faculty. Aleamoni and Hexner (1973) also investigated responses from students who were informed that their ratings, among other uses, would be considered in salary and promotion deliberations. These ratings were compared with students in a previous semester who were not told how the results would be used. Students who were told that the results would be used administratively rated the course more favorably.

In a more recent study (Centra, 1976a) found the differences in student ratings due to different written statements as to intended uses of the results were slight. One statement of intended use of the evaluation information mentioned salary, promotion, or tenure considerations for the teacher. The other statement said the information would be used only by the instructor to improve his or her teaching. He compared twenty-four items from the
evaluation instrument, SIR. Only five items yielded significant results. He concluded that students either did not read the special instructions carefully or simply did not respond more favorably when they were told the results would be used for administrative purposes.

Variables that influence student ratings have been widely researched as indicated in this review. Answers are not yet definitive enough to draw strong conclusions in most cases. Most of the variables can be controlled in analyses of data when identified. Current studies need to be continued particularly in those areas where conflicting results have been obtained. Student attitudes and variables which affect them can change with time as has been pointed out in this review. Research in the area of variables which affect student evaluations will have to continue to keep current in this far from static area.
CHAPTER THREE

Research Procedures

This chapter includes a description of the sample used in the study, the research procedures and a description of the instrument used for collection of the data.

Description of the Sample

The research was conducted at a private, post-secondary College of Business located in Upper East Tennessee. The school was accredited as a Senior College of Business by the Accrediting Commission of the Association of Independent Colleges and Schools (AICS). Approximately two-thirds of the students, however, were working toward Associate in Business Science Degrees in business areas. Among these curricular areas were accounting, banking and finance, business administration, computers, and office administration (legal, medical, and general).

A Bachelor of Science Degree was offered with majors in Business Administration or Office Administration. The school population was composed predominately of upper-middle and lower-middle socioeconomic groups. Over 75 percent of the students received some type of direct financial aid such as grants, loans, or college work study. The school had campuses in the three cities making up the
Tri-Cities of Upper East Tennessee. Approximately two hundred students were enrolled in each of the three centers. Twenty-six faculty members taught two or more classes during the Winter Quarter of 1983.

Ten faculty members were selected randomly from faculty that had a credit hour teaching load of at least six quarter hours. The random selection was made by drawing names of the ten instructors out of the proverbial hat in which the total eligible faculty names had been placed. Two classes taught by each of the ten instructors were selected randomly by the same method. Only classes that had at least ten students enrolled were considered eligible for the random selection.

**Procedures**

An experimental design was selected for the study. More specifically the design was identified by Borg and Gall (1979) as the posttest-only control-group design.

\[
\begin{align*}
R &\times X \\
R &\quad 0
\end{align*}
\]

\( R = \text{Random assignment} \)
\( X = \text{Experimental treatment} \)
\( O = \text{Posttest of the dependent variable} \)

The steps involved in this design are: (1) random assignment of the groups, (2) administration of the treatment to the experimental group but not to the control group, and (3) administration of the posttest to both groups (Borg
and Gall, 1979). This design was recommended by Borg when there is a possibility that a pretest may have an effect on the experimental treatment as could be the case in this study. The data yielded by this experimental design were analyzed by doing a t test comparison of the mean posttest scores of the experimental and the control groups.

The experimental design was further strengthened in this study because it was applied in an actual school setting. The groups were not artificially contrived but reflected actual classroom groups in their own environmental setting. Snow (1974:265) believed that educational researchers should design experiments to reflect the environment and the learner. That is, "experiments should become more representative of the natural environment and of human subjects as active learners."

All the faculty of the school were briefed by the researcher two weeks prior to the administration of the evaluation instruments to the classes (Appendix A). The instructor announced to the class one class period preceding the evaluation that an evaluation would be done the next class meeting by the researcher, that it would be anonymous, and that the results would in no way affect the grade for the class. All of the evaluation instruments were administered within four days. The classes evaluated were located on three campuses and consisted of both day and evening classes. The short time period during which
the evaluations were done as well as the times and places of the class meetings safeguarded discussion among students of the evaluation procedures.

The following procedure was uniformly used for administering the evaluation instruments to the two selected classes for each of the ten instructors. One of the classes for each instructor was randomly selected as the experimental group. In addition to the statement read to the control group about utilization, the experimental group was given the treatment variable emphasizing that the evaluation results would also be used for personnel decisions (Appendix B).

The researcher entered each classroom at the beginning of the class period with the SIR forms for that class. The instructor was given a cover sheet to mark for that class and asked to leave the room during the evaluation. The SIR forms were distributed to the class. Pencils (No. 2) were also distributed for marking. Instructions were given as to the proper way to mark the forms. Students were instructed that the forms would be collected in ten to fifteen minutes when all had completed the evaluation.

A prepared statement as to the utilization of the results was read (Appendix B). Students were asked if they had any questions as to marking and were then told to begin. Forms were collected by the researcher and
placed in the prepared envelope for that class. The teacher's cover sheet was attached to the envelope. The class was thanked for their participation and the researcher then left the room. The total time for instructions, marking the forms, collecting and distributing the forms was approximately twenty minutes.

Description of the Instrument

The Student Instructional Report (SIR) developed by the Educational Testing Service of Princeton, New Jersey, was chosen as the evaluation instrument for the study. Its selection as the evaluation instrument was based on: (1) the Student Instructional Report according to ETS typically and appropriately is used for instructional improvement and tenure, promotion, or salary decisions (Guidelines, 1981). (2) A continuing research program by ETS since 1972 supports the validity (Centra, 1976b) and reliability (Centra, 1973b) of the instrument. (3) Local scoring may be used for the instrument. (4) Institutional research is encouraged by ETS.

The short opinionnaire which requires ten to fifteen minutes to complete contained a total of thirty-nine items with space for the instructor to add additional items. For this study, no additional items were added. Thirty-one of the items involved students' ratings. The other eight items were largely descriptive information. The items
have also been factor analyzed into six major factors (Centra, 1973c). The factors identified by Centra's study and the SIR items in each category were:

**Factor I: Teacher-Student Relationship**

- 11 Student felt free to question or give opinions
- 19 Instructor openness to other viewpoints
- 7 Instructor encouraged students to think
- 10 Instructor raised challenging questions
- 8 Instructor concern with students' progress
- 4 Instructor availability for students
- 9 Instructor made helpful comments on papers or exams
- 5 Instructor knew when students didn't understand

**Factor II: Course Objectives and Organization**

- 2 Agreement between objectives and teaching
- 1 Course objectives made clear
- 20 Instructor accomplished objectives for the course
- 12 Instructor was well prepared for class
- 3 Instructor used class time well
- 13 Instructor informed students of how evaluated
- 14 Instructor summarized or emphasized major points

**Factor III: Lectures**

- 35 Overall rating of lectures
- 6 Lectures too repetitive of textbook(s)
- 39 Overall effectiveness of instructor
- 3 Instructor used class time well
- 16 Course scope was too limited
- 36 Overall value of class discussions

**Factor IV: Reading Assignments**

- 32 Overall rating of textbook(s)
- 33 Overall rating of readings
- 38 Overall value of course to student
- 15 Student interest stimulated by course

**Factor V: Course Difficulty and Workload**

- 21 Level of difficulty of the course
- 23 Pace of the course
- 22 Work load for the course

**Factor VI: Examinations**

- 34 Overall rating of exams
- 17 Exams reflected important aspects of the course
In the Factor Analysis Study done by Centra (1973c) the factors were reported to be fairly highly interrelated and were highly correlated with the global ratings of items 38 and 39. The first five factors were especially highly interrelated. Students who rated instructors high in one area of the SIR also tended to rate them high in others.

The six factors were given primary consideration in analysis of the data for this study. However, if the purpose of an evaluation study was to give actual feedback to instructors for improvement of instruction or to give actual feedback to the school for personnel decisions, an analysis of all the items and factors should be considered. The factors identify meaningful and useful clusters of items beyond that which might be accounted for by a single item. The separate factors do describe different aspects of instruction and it could well be that an instructor would be rated favorably in one area but unfavorably in another. Especially for the purpose of teaching improvement it would be most necessary to deal with each of the separate factors and their included items rather than a global score.
CHAPTER FOUR

Analysis of the Data

For an analysis of the data the hypotheses were stated in the null format:

$H_{01}$ Student evaluations of faculty using the mean scores from the SIR evaluation instrument will not show significant statistical differences at the .05 level between students that perceive the evaluation results will be used for improvement of instruction and students that perceive the evaluation results will be used for personal decisions.

$H_{02}$ Student evaluations of faculty using the mean scores from the SIR evaluation instrument will not show significant statistical differences at the .05 level between classes taught by full-time faculty and classes taught by part-time faculty.

$H_{03}$ Student evaluations of faculty using the mean scores from the SIR evaluation instrument will not show significant statistical differences at the .05 level between business or vocational classes and general studies classes.

$H_{04}$ Student evaluations of faculty using the mean scores from the SIR evaluation instrument will not show
significant statistical differences at the .05 level between classes taught by male teachers and classes taught by female teachers.

Testing of the Hypotheses

Means for the thirty-nine items on the SIR instrument for each of the twenty classes surveyed were computed using the Statistical Package for the Social Sciences (SPSS) at the computer center at East Tennessee State University. Data from the thirty-nine items on the instrument were grouped by class into six factor categories suggested by Centra (1973c) in a previous factor study of the SIR instrument. The factor categories used were as follows:

Factor 1 Teacher-Student Relationship
Factor 2 Course Objectives and Goals
Factor 3 Lectures
Factor 4 Reading Assignments
Factor 6 Examinations

The grouped means were then processed through a Commodore microcomputer to gain mean averages for each class in the six factor categories (Table 1). Class A was designated as the control group and Class B was designated as the experimental group. The total number of students surveyed was 303. SPSS was used with the results from these computations to perform several steps. The first was to perform t-tests for:
Table 1  
Summary of Teacher Means for Control (A) and Experimental (B) Group by Factor

<table>
<thead>
<tr>
<th>Teacher</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>27</td>
<td>2.915</td>
<td>3.081</td>
<td>2.984</td>
<td>3.150</td>
<td>3.387</td>
<td>3.139</td>
</tr>
<tr>
<td>A</td>
<td>9</td>
<td>3.408</td>
<td>3.660</td>
<td>3.285</td>
<td>4.189</td>
<td>2.889</td>
<td>3.188</td>
</tr>
<tr>
<td>A</td>
<td>19</td>
<td>3.382</td>
<td>3.698</td>
<td>3.552</td>
<td>4.091</td>
<td>3.685</td>
<td>3.917</td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>3.000</td>
<td>3.500</td>
<td>3.271</td>
<td>3.750</td>
<td>3.625</td>
<td>3.438</td>
</tr>
<tr>
<td>A</td>
<td>15</td>
<td>3.451</td>
<td>3.781</td>
<td>3.590</td>
<td>4.000</td>
<td>2.911</td>
<td>4.000</td>
</tr>
<tr>
<td>A</td>
<td>6</td>
<td>3.458</td>
<td>3.476</td>
<td>3.417</td>
<td>3.600</td>
<td>3.000</td>
<td>3.667</td>
</tr>
<tr>
<td>A</td>
<td>13</td>
<td>3.201</td>
<td>3.297</td>
<td>3.064</td>
<td>2.840</td>
<td>3.667</td>
<td>3.539</td>
</tr>
</tbody>
</table>
1. Control and experimental classes by factor one to factor six
2. Classes taught by male and female instructors by factor one to factor six
3. Classes taught by full-time and part-time faculty by factor one to factor six
4. Business or vocational classes and general studies classes by factor one to factor six.

Classes rather than individual student responses were used as the unit of analysis for the study.

**Comparison of the Control and Experimental Groups**

Mean difference scores between experimental and control groups for the classes taught by each of the ten instructors by the factors analyzed are presented in Table 2. These scores represent the mean differences between the class which was told that the evaluation results would be used for improvement of instruction (control group) and the class which was told that the evaluation results would be used additionally for input for personnel decisions (experimental group) for each instructor. Positive differences show higher ratings for the experimental classes. The summarized mean differences for each of the factors, however, in fact, show differences in a negative direction on all but one factor, Factor 4, which deals with reading.
Table 2
Differences in the Means for Experimental and Control Classes for Each Instructor

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Factor 1: Teacher-Student Relationship</th>
<th>Factor 2: Objectives and Goals</th>
<th>Factor 3: Lectures</th>
<th>Factor 4: Reading Assignment</th>
<th>Factor 5: Course Workload</th>
<th>Factor 6: Examinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.283</td>
<td>0.280</td>
<td>0.294</td>
<td>0.462</td>
<td>-0.054</td>
<td>0.236</td>
</tr>
<tr>
<td>2</td>
<td>-0.272</td>
<td>-0.151</td>
<td>-0.181</td>
<td>-0.619</td>
<td>0.194</td>
<td>-0.069</td>
</tr>
<tr>
<td>3</td>
<td>0.005</td>
<td>-0.054</td>
<td>-0.078</td>
<td>0.109</td>
<td>-0.139</td>
<td>0.291</td>
</tr>
<tr>
<td>4</td>
<td>0.036</td>
<td>0.101</td>
<td>-0.051</td>
<td>-0.014</td>
<td>-0.100</td>
<td>0.037</td>
</tr>
<tr>
<td>5</td>
<td>-0.382</td>
<td>-0.198</td>
<td>-0.281</td>
<td>-0.341</td>
<td>-0.060</td>
<td>-0.479</td>
</tr>
<tr>
<td>6</td>
<td>-0.335</td>
<td>-0.081</td>
<td>-0.066</td>
<td>-0.215</td>
<td>-0.381</td>
<td>-0.018</td>
</tr>
<tr>
<td>7</td>
<td>0.065</td>
<td>-0.313</td>
<td>-0.238</td>
<td>-0.228</td>
<td>0.163</td>
<td>-0.388</td>
</tr>
<tr>
<td>8</td>
<td>-0.043</td>
<td>-0.042</td>
<td>-0.079</td>
<td>-0.059</td>
<td>0.039</td>
<td>-0.023</td>
</tr>
<tr>
<td>9</td>
<td>0.076</td>
<td>0.101</td>
<td>0.085</td>
<td>0.686</td>
<td>0.071</td>
<td>0.104</td>
</tr>
<tr>
<td>10</td>
<td>0.193</td>
<td>0.300</td>
<td>0.341</td>
<td>0.998</td>
<td>-0.042</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Factor Totals: -0.037 -0.005 -0.025 0.077 -0.030 -0.006
assignments. The greatest differences, all in a negative direction, are noted for Factor 1 dealing with teacher-student relationships with a summarized mean difference of -0.037, Factor 3 dealing with lectures with a summarized mean difference of -0.025, and Factor 5 dealing with course workload with a summarized mean difference of -0.030.

Summarized mean scores for all experimental and control classes and their t values are presented in Table 3. The data in Table 3 indicate that the classes did not differ significantly on any of the factor groupings at the .05 level of significance.

When the factor categories failed to show significant differences for the experimental and control classes, a more detailed look at the global rating items, 38 and 39, was conducted. These global rating items are the ones usually consulted when an overall look at the evaluation is needed. Teacher means for SIR Item 38 are presented in Table 4. Item 38 deals with the overall value of the course to the student. For teachers 2, 3, 5, 6, 7, or one-half of the teachers involved in this study, the mean for the control class was higher than the mean for the experimental class but not at a significant level. Four of these instructors were full-time teachers; three were male instructors. The two classes for each instructor were in the same subject area. No consistent pattern for this difference in a negative direction was evident.


<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Classes</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>t-value</th>
<th>Degrees of Freedom</th>
<th>2-tail Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher-Student Relationship</td>
<td>Control</td>
<td>10</td>
<td>3.3044</td>
<td>0.180</td>
<td>18</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>10</td>
<td>3.2667</td>
<td>0.190</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Objectives and Goals</td>
<td>Control</td>
<td>10</td>
<td>3.4522</td>
<td>0.232</td>
<td>18</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>10</td>
<td>3.4474</td>
<td>0.113</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lectures</td>
<td>Control</td>
<td>10</td>
<td>3.2894</td>
<td>0.213</td>
<td>18</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>10</td>
<td>3.2641</td>
<td>0.130</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Assignments</td>
<td>Control</td>
<td>10</td>
<td>3.6021</td>
<td>0.510</td>
<td>18</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>10</td>
<td>3.6800</td>
<td>0.105</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Workload</td>
<td>Control</td>
<td>10</td>
<td>3.1683</td>
<td>0.325</td>
<td>18</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>10</td>
<td>3.1379</td>
<td>0.344</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examinations</td>
<td>Control</td>
<td>10</td>
<td>3.5704</td>
<td>0.277</td>
<td>18</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>10</td>
<td>3.5641</td>
<td>0.220</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Table 4
Teacher Means for SIR Item 38 by Class

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Variable</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td></td>
<td>Teacher 6</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>3.550</td>
<td>Experimental</td>
<td>4.000</td>
</tr>
<tr>
<td>Control</td>
<td>3.391</td>
<td>Control</td>
<td>4.313</td>
</tr>
<tr>
<td>Teacher 2</td>
<td></td>
<td>Teacher 7</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.125</td>
<td>Experimental</td>
<td>4.235</td>
</tr>
<tr>
<td>Control</td>
<td>4.667</td>
<td>Control</td>
<td>4.600</td>
</tr>
<tr>
<td>Teacher 3</td>
<td></td>
<td>Teacher 8</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.000</td>
<td>Experimental</td>
<td>4.176</td>
</tr>
<tr>
<td>Control</td>
<td>4.250</td>
<td>Control</td>
<td>4.000</td>
</tr>
<tr>
<td>Teacher 4</td>
<td></td>
<td>Teacher 9</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.188</td>
<td>Experimental</td>
<td>3.833</td>
</tr>
<tr>
<td>Control</td>
<td>4.091</td>
<td>Control</td>
<td>3.154</td>
</tr>
<tr>
<td>Teacher 5</td>
<td></td>
<td>Teacher 10</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.000</td>
<td>Experimental</td>
<td>4.286</td>
</tr>
<tr>
<td>Control</td>
<td>4.529</td>
<td>Control</td>
<td>3.688</td>
</tr>
</tbody>
</table>
The greatest difference was noted in teachers 9 and 10 with higher mean scores in the experimental group and mean differences of .679 and .602 respectively. The control class for each of these instructors was in a subject area outside their regular teaching area. Teacher 9's control class was office filing and her teaching area was history. Teacher 10's control class was an introduction to computers while his regular teaching area was mathematics. The instructor teaching outside his field could have been a factor in this greater difference rather than the student's perception of the use of the evaluation results. None of the mean differences for Item 38 were significant at the .05 level.

Mean data for Item 39, a global rating item for overall effectiveness of the instructor, were looked at for the experimental and control class for each instructor (Table 5). Again, no consistent pattern was evident. Teacher 1 and Teacher 10 had mean differences of .750 and 1.167 respectively in a positive direction. As has been pointed out earlier, Teacher 10's control class was in a subject area outside his regular teaching field. This was not true for Teacher 1. Both of these teachers were male full-time teachers. These differences were not significant at the .05 level and for Teacher 10 the greater difference might be attributed to the subject area of the control class.
Table 5

Teacher Means for SIR Item 39 by Class

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Variable</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td></td>
<td>Teacher 6</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.333</td>
<td>Experimental</td>
<td>4.333</td>
</tr>
<tr>
<td>Control</td>
<td>3.583</td>
<td>Control</td>
<td>4.313</td>
</tr>
<tr>
<td>Teacher 2</td>
<td></td>
<td>Teacher 7</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.000</td>
<td>Experimental</td>
<td>4.278</td>
</tr>
<tr>
<td>Control</td>
<td>4.556</td>
<td>Control</td>
<td>4.833</td>
</tr>
<tr>
<td>Teacher 3</td>
<td></td>
<td>Teacher 8</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>3.636</td>
<td>Experimental</td>
<td>4.471</td>
</tr>
<tr>
<td>Control</td>
<td>3.933</td>
<td>Control</td>
<td>4.667</td>
</tr>
<tr>
<td>Teacher 4</td>
<td></td>
<td>Teacher 9</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>4.600</td>
<td>Experimental</td>
<td>3.538</td>
</tr>
<tr>
<td>Control</td>
<td>4.600</td>
<td>Control</td>
<td>3.615</td>
</tr>
<tr>
<td>Teacher 5</td>
<td></td>
<td>Teacher 10</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>3.875</td>
<td>Experimental</td>
<td>4.667</td>
</tr>
<tr>
<td>Control</td>
<td>4.813</td>
<td>Control</td>
<td>3.500</td>
</tr>
</tbody>
</table>
For Teacher 4 the mean for control and experimental class was the same. Differences for the other seven teachers were all in a negative direction with the control class having the higher mean. The greatest negative differences were noted for Teacher 2 and Teacher 5 with mean differences of 0.556 and -1.038 respectively. Both of the teachers were full-time and teaching in their subject area. One was a female instructor and the other was a male instructor. None of the differences for Item 39 were significant at the .05 level.

On the basis of these data, null hypothesis 1 concerning student perception of the use of the evaluation results was accepted as stated.

**Comparison of the Classes Taught by Full-time Faculty and Classes Taught by Part-time Faculty**

Summarized mean scores for classes taught by full-time faculty and classes taught by part-time faculty are presented in Table 6. It may be noted that part-time faculty classes had a higher mean for Factor 1 dealing with teacher-student relationships, Factor 2 dealing with course objectives and goals, Factor 3 dealing with lectures, and Factor 6 dealing with examinations. The t values (Table 6) indicate that none of the differences were significant at the .05 level. However, the t value of -1.82 for Factor 1, teacher-student relationships, with a 2-tail probability of 0.086, while not statistically significant at the .05 level,
<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Classes</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>t-value</th>
<th>Degrees of Freedom</th>
<th>2-tail Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1: Teacher-Student Relationship</td>
<td>Full-time</td>
<td>14</td>
<td>3.2400</td>
<td>0.186</td>
<td>-1.82</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Part-time</td>
<td>6</td>
<td>3.3918</td>
<td>0.124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 2: Course Objectives and Goals</td>
<td>Full-time</td>
<td>14</td>
<td>3.4242</td>
<td>0.181</td>
<td>-0.98</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Part-time</td>
<td>6</td>
<td>3.5095</td>
<td>0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 3: Lectures</td>
<td>Full-time</td>
<td>14</td>
<td>3.2508</td>
<td>0.161</td>
<td>-1.03</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Part-time</td>
<td>6</td>
<td>3.3373</td>
<td>0.197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 4: Reading Assignments</td>
<td>Full-time</td>
<td>14</td>
<td>3.6574</td>
<td>0.356</td>
<td>0.30</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Part-time</td>
<td>6</td>
<td>3.6030</td>
<td>0.403</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 5: Course Workload</td>
<td>Full-time</td>
<td>14</td>
<td>3.1569</td>
<td>0.278</td>
<td>0.08</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Part-time</td>
<td>6</td>
<td>3.1443</td>
<td>0.452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 6: Examinations</td>
<td>Full-time</td>
<td>14</td>
<td>3.5107</td>
<td>0.068</td>
<td>-1.66</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Part-time</td>
<td>6</td>
<td>3.6992</td>
<td>0.065</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
is noteworthy as the greatest t value obtained in the factor comparisons. On the basis of the data presented, null hypothesis 2 concerning classes taught by full-time and part-time faculty was accepted.

Comparison of Business or Vocational Classes and General Studies Classes

Summarized mean scores for business or vocational classes and general studies classes and their t values are presented in Table 7. General studies classes had higher mean averages for all the factor groupings except Factor 4, reading assignments, where the mean difference was -0.081 and the t value was -.49. The t values did not indicate differences significant at the .05 level; therefore, null hypothesis 3 concerning business or vocational classes and general studies classes was accepted.

Comparison of the Classes Taught by Male Teachers and the Classes Taught by Female Teachers

Summarized mean scores for the classes taught by male teachers and the classes taught by female teachers and their resultant t values are presented in Table 8. The total mean is higher in every case for classes taught by male teachers. No differences were significant at the .05 level as revealed by the t values. However, Factor 3 dealing with lectures did show the greatest mean difference, 0.2406, with a resultant t value of 1.91 and 2-tail probability of 0.072. On the basis of the data presented,
Table 7

$t$ Scores for Business or Vocational and General Studies Classes by Factor

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Classes</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>$t$-value</th>
<th>Degrees of Freedom</th>
<th>2-tail Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher-Student</td>
<td>Gen. Studies</td>
<td>10</td>
<td>3.3271</td>
<td>0.214</td>
<td>1.03</td>
<td>18</td>
</tr>
<tr>
<td>Relationship</td>
<td>Bus. or Voc.</td>
<td>10</td>
<td>3.2440</td>
<td>0.141</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Objectives and Goals</td>
<td>Gen. Studies</td>
<td>10</td>
<td>3.4582</td>
<td>0.169</td>
<td>0.21</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Bus. or Voc.</td>
<td>10</td>
<td>3.4414</td>
<td>0.195</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lectures</td>
<td>Gen. Studies</td>
<td>10</td>
<td>3.3292</td>
<td>0.179</td>
<td>1.40</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Bus. or Voc.</td>
<td>10</td>
<td>3.2243</td>
<td>0.157</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Assignments</td>
<td>Gen. Studies</td>
<td>10</td>
<td>3.6004</td>
<td>0.357</td>
<td>-0.49</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Bus. or Voc.</td>
<td>10</td>
<td>3.6817</td>
<td>0.379</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Workload</td>
<td>Gen. Studies</td>
<td>10</td>
<td>3.2120</td>
<td>0.362</td>
<td>0.80</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Bus. or Voc.</td>
<td>10</td>
<td>3.0942</td>
<td>0.293</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examinations</td>
<td>Gen. Studies</td>
<td>10</td>
<td>3.6260</td>
<td>0.070</td>
<td>1.08</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Bus. or Voc.</td>
<td>10</td>
<td>3.5085</td>
<td>0.083</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8

$t$ Scores for Classes Taught by Male and Female Teachers by Factor

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>Number of Classes</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>t-value</th>
<th>Degrees of Freedom</th>
<th>2-tail Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher-Student Relationship</td>
<td>Male</td>
<td>12</td>
<td>3.3002</td>
<td>0.206</td>
<td>0.43</td>
<td>18</td>
<td>0.669</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8</td>
<td>3.2635</td>
<td>0.147</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Objectives and Goals</td>
<td>Male</td>
<td>12</td>
<td>3.4829</td>
<td>0.191</td>
<td>1.02</td>
<td>18</td>
<td>0.320</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8</td>
<td>3.4001</td>
<td>0.153</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lectures</td>
<td>Male</td>
<td>12</td>
<td>3.3330</td>
<td>0.175</td>
<td>1.91</td>
<td>18</td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8</td>
<td>3.1924</td>
<td>0.136</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Assignments</td>
<td>Male</td>
<td>12</td>
<td>3.6747</td>
<td>0.365</td>
<td>0.50</td>
<td>18</td>
<td>0.623</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8</td>
<td>3.5906</td>
<td>0.373</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Workload</td>
<td>Male</td>
<td>12</td>
<td>3.1630</td>
<td>0.317</td>
<td>0.16</td>
<td>18</td>
<td>0.873</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8</td>
<td>3.1382</td>
<td>0.361</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examinations</td>
<td>Male</td>
<td>12</td>
<td>3.6075</td>
<td>0.250</td>
<td>0.90</td>
<td>18</td>
<td>0.380</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>8</td>
<td>3.5069</td>
<td>0.237</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
null hypothesis 4 concerning male and female instructors was accepted.

Discussion of Results

The differences in student ratings due to different instructions as to intended uses of the results were slight. Data from earlier research led to the prediction that ratings would be higher when students were informed that the results would be used as input for personnel decisions. Two studies done in the 1950's (Wherry, 1952; Taylor and Wherry, 1951) reported convincing evidence that raters evaluated instructors higher when they were informed the results would be used for administrative purposes. Aleamoni and Hexner (1973) also reported higher ratings from students who were informed their ratings would be considered in salary and promotion deliberations.

The data in this study, however, showed no significant differences in any of the factor groupings between experimental and control classes. Global rating items dealing with overall value of the class and overall teacher effectiveness also failed to show significant differences for experimental and control classes. These global rating items would probably be given special attention if the ratings were used for personnel decisions. The instructor in several classes received higher mean ratings from the control group in the factor categories as well as the overall rating items.
Even though no significant differences were obtained when classes taught by part-time faculty were compared to classes taught by full-time faculty, part-time faculty received higher mean ratings for Factor 1, teacher-student relationship, as well as for three of the other factor groupings. An earlier study (Centra, 1976b) showed the lowest ratings for teachers with four to six hour loads and the highest ratings for teachers with teaching loads of thirteen or more hours. Only six classes in the study were taught by part-time faculty.

Although no previous studies were brought to light that compared business or vocational classes and general studies classes, the results of this study were not conclusive. No significant differences emerged from comparisons in any of the factor groupings; however, higher mean averages in general studies classes were obtained in all the factor groupings except one. These results were consistent with previous research studies showing higher student ratings in the humanities (Centra and Creech, 1976; Feldman, 1978; Centra, 1972). Ten classes in this study were general studies classes and ten were business or vocational classes. The general studies classes surveyed included economics, government, anatomy, English, math, and statistics. The business and vocational classes surveyed included typing, office machines, business filing, accounting, income tax, business law, principles of investment, and introduction to computers.
Conflicting results have been obtained from previous studies when relating sex of the instructor to students' evaluations of instruction. The results of this study did not clarify the conflicting results of previous studies at a statistically significant level; however, male instructors did receive higher mean scores in all the factor categories. Factor 3 dealing with lectures showed the greatest mean difference. Eight classes taught by females and twelve classes taught by males were surveyed for the data analyzed in this study.
CHAPTER FIVE

Summary, Conclusions and Implications

Summary

Data for this study were collected from twenty classes taught by ten instructors at a private post-secondary institution in East Tennessee. Both instructors and their classes were randomly selected. The total number of students surveyed was 303.

The Student Instructional Report (SIR) was used as the instrument for collection of the student data. Two classes for each of the ten instructors were surveyed by the investigator. The control group was given oral instructions and information to the effect that the results of the evaluation survey would be used for the improvement of instruction. The experimental group was given the same oral instruction and information with the additional treatment that the results would have input for personnel decisions (Appendix B).

Data from the thirty-nine items on the SIR instruments were grouped by class into six factor categories suggested by Centra (1973c) in a previous factor study of the SIR instrument. The factor categories were as follows:
Factor 1. Teacher-Student Relationships
Factor 2. Course Objectives and Goals
Factor 3. Lectures
Factor 4. Reading Assignments
Factor 5. Course Workload
Factor 6. Examinations

Means for the thirty-nine items for each of the twenty classes were computed using the Statistical Package for the Social Sciences (SPSS). Grouped means in the six factor categories were used to perform t-tests for (1) control and experimental classes by Factor 1 to Factor 6, (2) classes taught by male and female instructors by Factor 1 to Factor 6, (3) classes taught by full-time and part-time faculty by Factor 1 to Factor 6, (4) business or vocational classes and general studies classes by Factor 1 to Factor 6. Classes were used as the unit of analysis for the study.

Differences in student ratings due to different instructions on the intended uses of the results were not statistically significant at the .05 level of significance. Data were analyzed for experimental and control classes for each of the factor categories as well as for global rating items 38 and 39. No consistent pattern emerged in the data analysis. Although not at a statistically significant level, negative differences indicating the control classes
had a higher summarized mean occurred for five factors in a summary of the mean scores of the six factor categories. Five teachers received higher mean scores for the control class for item 38 dealing with overall value of the class to the student. Six teachers received higher mean scores for the control class for item 39 dealing with overall teacher effectiveness.

Summarized data from all classes taught by part-time faculty showed higher mean scores in four of the factor categories. Factor 1, teacher-student relationship, showed the highest mean for part-time faculty and the greatest difference in the t comparisons with full-time faculty. None of the differences from the comparison of full-time and part-time faculty were significant at the .05 level in any of the six factor categories.

When summarized mean scores for business or vocational classes and general studies classes were compared, t values did not indicate differences significant at the .05 level. General studies classes did receive higher mean scores for all the factor categories except Factor 4, dealing with reading assignments.

Male teachers received higher mean scores for all the factor categories based on summarized mean scores for all their classes. When compared to the summarized mean scores from classes taught by female instructors, the
differences were not significant at the .05 level. However, Factor 3 dealing with lectures did show the greatest mean difference between classes taught by male and female instructors.

The data from the study seem to indicate that the students at the institution surveyed were discriminating and relatively free from the influences of the variables in this study in their evaluative ratings of faculty and instruction.

Conclusions

As a result of this study, using the class mean scores from the Student Instructional Report (SIR) evaluation instrument and t test for statistical analysis, the following conclusions were drawn.

1. There were no statistically significant differences at the .05 level between student groups that were told the evaluation results would be used for improvement of instruction and student groups that were told the evaluation results would have input for personnel decisions.

2. There were no statistically significant differences at the .05 level based on the evaluation results between student groups taught by full-time faculty and student groups taught by part-time faculty.
3. There were no statistically significant differences at the .05 level based on the evaluation results between business or vocational classes and general studies classes.

4. There were no statistically significant differences at the .05 level based on the evaluation results between classes taught by male instructors and classes taught by female instructors.

Implications

Several implications may help to explain the lack of statistically significant results in this study. Since the 1950's when earlier studies were done on this same premise, student evaluations of faculty have been increasingly used as measures of teaching effectiveness. Students may be becoming more discriminating and less susceptible to the variables of this study as they become more sophisticated about evaluation instruments. Also, as Centra (1976a:282) observed in his study, "Another possibility, however, is that many students simply did not respond more leniently when they were told the results would be used for administrative purposes." Hopefully, this could indicate that students are taking their job of evaluating teachers seriously enough to not be unduly influenced in their task whatever the use of the evaluation results.

It may be cautiously surmised from the present study that student raters at the institution surveyed were in general free from the variable influences investigated in
this study at least at a statistically significant level. Careful analysis of the data did suggest that effects on student evaluations may have occurred but the differences in ratings did not appear large enough or consistent enough to have significance.

Future studies repeating the evaluations with the same instructors through follow-up terms might yield additional and more conclusive results. Additional research involving part-time and full-time instructors should be encouraged because of the lack of evaluation studies in this specific area.
REFERENCES
REFERENCES

Aleamoni, L. M. and P. Z. Hexner

Aleamoni, L. M. and P. Z. Hexner

Aleamoni, L. M. and G. S. Thomas
1977 "Is the instructor's rating of the class related to the class' rating of the instructor?" Research Report No. 1. Tucson: University of Arizona, Office of Instructional Research and Development.

Aleamoni, L. M. and M. Yimer

Alkin, M. C., R. Daillak, and P. White

Aubrecht, Judith D.

Bogue, E. Grady

Boozer, Charles H., Mickey M. Lee, and Jack H. Rayson
Borg, Walter R. and Meredith D. Gall  
1979 Educational Research, An Introduction. New York:  
Longman, Inc.

Carlson, Keith W. and Patricia G. Bensinger  
1981 "Multiple Correlational Analysis of Professor's  
Ratings and Selected Student Variables: a pre­  
liminary study." The Educational Catalyst.  
Middle Tennessee State University: Divisions of  
Teacher Education in Tennessee Public Institutions  
of Higher Learning, XI, 1 (Spring).

Carlson, P.  
1974 "The relationship of three organizational variables  
on the utilization of evaluation." Doctoral dis­  
sertation, Ohio State University, Dissertation  
Abstracts International, 34 7097A (University  
Microfilms No. 74-10):924.

Cashin, William E.  
1978 Sources of information upon which to base evalu­  
ations of college teaching. Manhattan, Kansas:  
Center for Faculty Evaluation and Development in  
Higher Education, Kansas State University.

Centra, John A.  
1964 Michigan State University Student Instructional  
Michigan State University: Office of Institutional  
Research.

Centra, John A.  
1972 The Utility of Student Ratings for Instructional  
Improvement. Princeton, N. J.: Educational  
Testing Service.

Centra, John A.  
1973a "The relationship between student and alumni ratings  
of teachers." SIR Report No. 3. Princeton:  
Educational Testing Service.

Centra, John A.  
1973b "Reliability of the student instructional report  
items." SIR Report No. 3, Princeton: Educational  
Testing Service.

Centra, John A.  
1973c "Factor analysis of the student instructional  
report items." SIR Report No. 3. Princeton:  
Educational Testing Service.
Centra, John A.

Centra, John A.

Centra, John A.

Centra, John A. and F. Reid Creech

Costin, F., W. T. Greenough, and R. J. Menges

Costin, F. and J. E. Grush

Dickey, B.

Doyle, K. O. Jr., and S. E. Whitely

Drucker, A. T. and H. H. Remmers
1950 "Do alumni and students differ in their attitude toward instructors?" Purdue University Studies of Higher Education 70: 62-74.
Elliot, D. W.
1950 "Characteristics and relationships of various criteria of college and university teaching."
Purdue University Studies of Higher Education 70:5-61.

Feldman, K. A.

Feldman, K. A.

Feldman, K. A.
1979 "The significance of circumstances for college students' rating of their teachers and courses." Research in Higher Education 10:149-172.


Guthrie, D. R.

Haenn, J. F.
1980 "Reasons why evaluations and testing don't inform."

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

Kaercher, Dan
1882 "Teacher competency, how well are our kids being taught?" Better Homes and Gardens 60, 9 (September):19-24.

Kirkland, Tom
1982 "Peer pressure cited as teaching motivator."
Kronk, Annie K. and Thomas A. Shipka
1980 Evaluation of Faculty in Higher Education.
Washington, D. C.: National Education
Association.

Kulik, J. A. and W. J. McKeachie
1975 "The evaluation of teachers in higher education."
Review of Research in Education, Itasca, IL:
F. E. Peacock.

McClellan, Stephanie (Ed.)
1982 "Faculty evaluations should be revised." East
Tennessean LVII, 14 (October):28.

McKeachie, W. J. and Y. Lin
1971 "Sex differences in student responses to college
teachers: teacher warmth and teacher sex."

Meltsner, A. J.
1976 Policy Analysts in the Bureaucracy. Berkeley:
University of California Press.

Millman, Jason (Ed.)
1981 Handbook of Teacher Evaluation. Beverly Hills:
Sage Publications.

Miller, R. I.
1972 Evaluating Faculty Performance. San Francisco:

1980 "Communication theory and the utilization of
(Eds.) Utilization of Evaluation Information.
San Francisco: Jossey-Bass:30-33.

Patton, M. Q., B. S. Grimes, K. M. Guthrie, N. J. Brennan,
B. D. Grench, and D. A. Blyth.
1978 "In search of impact: An analysis of the utiliza­
tion of federal health evaluation research."
T. D. Cook and Associates (Eds.), Evaluation
Studies Review Annual. Volume 3, Beverly Hills:
Sage Publications. 141-150.

Perry, Suzanne
1982 "Fair and useful evaluation of professors by
students: experts say it can be done." The
Chronicle of Higher Education XXV, 14 (December).


Remmers, H. H. 1929 "The college professor as the student sees him." Purdue University Studies of Higher Education 29:75.


SIR Comparative Data Guide. Princeton: Educational 1979 Testing Service


Stake, Robert E.

Taylor, E. K. and R. J. Wherry

Thorne, G. S., C. S. Scott, and J. J. Beaird

Toch, Thomas
1982 "Miami union cooperates on tough evaluations." Education Week II, 13 (December).

University of South Florida Course and Instructor Form.
1965 Unpublished mimeograph.

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

Ware, John E. and Reed G. Williams

Weeks, E. C.

Weiner, S. S., D. Rubin, and T. Sachse

Weiss, C. H. (Ed.)
Wherry, R. J.

Whitely, Susan E. and Kenneth O. Doyle

Williams, Reed C. and John E. Ware, Jr.
APPENDICES
APPENDIX A

OVERVIEW REMARKS OF MEETING WITH BRISTOL COLLEGE AND HANDOUT TO FACULTY
Overview Remarks

1. Ten instructors at Bristol College have been chosen randomly to participate in an evaluation research study as part of my dissertation efforts.

2. Two classes for each of the instructors have also been selected randomly to be evaluated.

3. The Student Instructional Report (SIR) has been selected as the evaluation instrument. This instrument was developed by Educational Testing Service, Princeton, New Jersey.

4. One of the classes selected for each instructor will be told that the evaluation will be used for improvement of instruction. The second class will be told in addition that the evaluation will be used for personnel decisions.

5. Summarized data from these two groups will be compared statistically to determine whether there are significant differences in the evaluation results.

6. Both students and teachers will remain anonymous in reporting the evaluation results.

7. Bristol College will receive summaries of the evaluative data.

8. The above information is for your enlightenment. Please do not discuss this with your students as it would have a biasing effect on the study.
Handout to Faculty

Special Instructions to Faculty of Bristol College
for Administration of Evaluation

1. Distribution of sample copies of SIR to faculty with explanation

2. Explanation of cover sheet for faculty

3. Explanation of procedure for collection of data

   a. You will be notified by the Dean after this meeting if you have been selected for the study

   b. You will be contacted then, by me next week about the classes to be evaluated and the time. Evaluations will be conducted the week of Jan. 24-27.

   c. I will come to your class at the beginning of the period, give you the cover sheet, and ask you to leave the room while students mark the evaluation forms.

   d. After the forms are collected you will be notified so that you can resume your class. The total procedure should last 15-20 minutes.

   e. Please notify each class chosen to be evaluated, after we have discussed time, etc., one class period before the evaluation that a student evaluation will be conducted next class meeting by me. Please emphasize that they will not sign the evaluation form and that it will in no way affect their grade.
APPENDIX B

EXPLANATIONS TO CONTROL AND EXPERIMENTAL GROUPS

A. **Utilization Statement for Control Group**

We want your input for the improvement of instruction at Bristol College. One way for us to get this is to ask you to complete this evaluation form. Answering each question accurately and honestly will provide the most helpful information.

B. **Utilization Statement for Experimental Group**

We want your input for the improvement of instruction at Bristol College. As student raters, you should also know that the results of your ratings will be used as part of the information considered by Bristol College to make personnel decisions. Decisions such as which instructors should be promoted or given pay raises are important decisions. Please be both accurate and honest when you answer each question.
APPENDIX C

PERMISSION LETTERS FROM EDUCATIONAL TESTING SERVICE
AND BRISTOL COLLEGE
January 11, 1983

Mr. David M. Tiffany
Associate Vice President
Academic Affairs
East Tennessee State University
Box 24480A
Johnson City, TN 37614-0002

Dear Mr. Tiffany:

Thank you for your letter concerning use of the Student Instructional Report by one of your graduate students in her dissertation research. We have no objection to her use of 200-300 SIR answer sheets at Bristol College for that study.

We did a similar study with SIR several years ago. Your student will want to look at the report of the results: "The Influence of Different Directions on Student Ratings of Instruction." ETS Research Bulletin 75-28. It also was published in the Journal of Educational Measurement, Winter 1976 which she probably can find in the library.

If she has any questions as she proceeds, she should feel free to write or call me.

Sincerely yours,

Nancy Beck
Program Director

NB/em

77
January 24, 1983

Dr. David N. Tiffany
Associate Vice President
Academic Affairs
East Tennessee State University
Box 24490A
Johnson City, TN 37601

Dear Dr. Tiffany:

Bristol College is pleased to participate with Ann James in an evaluation study of our faculty. We realize the prime importance of teacher evaluation as we strive to keep our instruction and curriculums of the highest caliber.

Mrs. James will have the support of our faculty and administration when she comes to the college January 24-27, 1983, to administer the SIR evaluations to our students.

Cordially,

Jack O. Anderson, Ed. D.
President

JOA/rad

cc: Ann James
VITA

ANN NEBLET JAMES

Personal Data:
Place of Birth: Coeburn, Virginia
Current Residence:
1033 Grace Drive
Johnson City, Tennessee

Education:
East Tennessee State University,
Johnson City, Tennessee; Educational
Supervision and Administration, Ed.D.
East Tennessee State University,
Johnson City, Tennessee; Science
Education, M. A.
University of Richmond, Richmond,
Virginia; Biology, Chemistry, B. S.
Virginia Intermont College, Bristol,
Virginia; A. A. Liberal Arts.

Professional Experience:
Steed College, Johnson City, Tennessee;
Academic Dean.
Steed College, Johnson City, Tennessee;
Director, Medical Administrative Assistant
Programs.
Johnson City School System, Central
Office, Student Intern.
Steed College, Johnson City, Tennessee;
Chairman, Office Administration Division
Steed College, Johnson City, Tennessee;
Chairman, Medical Secretarial Department
University High, East Tennessee State
University, Johnson City, Tennessee;
Teacher, Biology and Science.
Bristol, Virginia School System, Bristol,
Virginia; Teacher, Biology, H. S.
Wise County School System, Wise, Virginia;
Teacher, 7th grade science
Center for Disease Control, Atlanta,
Georgia; Bacteriologist.
Georgia State Department of Health,
Atlanta, Georgia; Microscopist.
City of Richmond, Department of Public
Health, Richmond, Virginia; Student Intern
Professional Organizations:
American Association for the Advancement of Science
Phi Delta Kappa
Beta Beta Beta

Publications:

Honors:
Doctoral Fellow, ETSU, Department of Supervision and Administration, Supervision of Student Teachers.
Doctoral Fellow, ETSU, Department of Supervision and Administration, Research and Development. Editor of *The Communicator*. 