AN ABSTRACT OF THE THESIS OF

Daniel Anstine for the degree of Doctor of Philosophy in Education presented on September 12, 1991.

Title: An Investigation of the Relationships Between Student Evaluations and Faculty, Class, and Student Demographic Variables in Rating Instructional Effectiveness

Abstract approved: Redacted for Privacy
Robert Rice

The purpose of this study was to analyze and interpret the results of fall term student ratings of faculty teaching performance (n= 40,000) in order to determine whether relationships exist between a range of demographic variables and the ratings professors receive from students enrolled in their classes. A population of 645 full-time faculty and 273 teaching assistants from 11 colleges across Oregon State University were evaluated in fall term 1989-1990 by students who completed the university's Student Assessment of Teaching Instrument (SATI).

The results from an analysis of demographic variables and faculty ratings revealed that students rated faculty significantly different in major and non-major courses. No differences were found between faculty ratings and student variables of GPA, expected grade, and percent of classroom attendance.

Overall faculty ratings in large and small classes were not significantly different, although differences were found on several SATI instructional dimensions. Students in large classes rated faculty higher

on clearly presenting course objectives/requirements, being well prepared and organized, clearly presenting subject matter concepts, and relevancy of examinations. Faculty in small classes received the highest ratings on the sensitivity of the instructor to student understanding material presented in class, or having more opportunity for student/faculty interaction.

Statistical differences were found between graduate and undergraduate ratings of faculty teaching performance. Freshman class level in undergraduate courses and graduate students gave significantly higher faculty ratings than other levels. Significant gender differences were also found in student ratings of faculty in the Colleges of Engineering and Home Economics. Female faculty in Engineering received significantly higher ratings than male faculty by male students. Conversely, male faculty in Home Economics received significantly higher ratings than female faculty by female students.

There was no significant difference between tenure status and faculty ratings; however a difference did appear when tenure was compared with faculty productivity rates (publications in refereed journals). No significant differences were found between publication rates and faculty rank. Differences in faculty publication rates appeared when gender of the faculty was analyzed.

There was no significant difference between teaching performance ratings for all academic ranks of full-time faculty and non-international teaching assistants (NITAs). Statistical differences were noted for international teaching assistants (ITAs) and all academic ranks. Significant difference were found between NITAs and ITAs overall ratings as well as along several instructional dimensions of the SATI.

An Investigation of the Relationships Between Student Evaluations and Faculty, Class, and Student Demographic Variables in Rating Instructional Effectiveness

by

Daniel D. Anstine

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed September 12, 1991

Commencement June 1992

APPROVED:

Redacted for Privacy

Assistant Professor of Postsecondary and Technological Education in charge of major

Redacted for Privacy

Chair of the Department of Postsecondary and Technological Education

Redacted for Privacy

Dean of the College of Education

Redacted for Privacy

Dean of Graduate School

Date thesis is presented September 12, 1991

Typed by D. Anstine for Daniel Anstine

to

My wife Jiine and daughter Sarah

who endured and sacrificed much while

this work was being completed

ACKNOWLEDGEMENT

The author is very grateful to Dr. Robert Rice for serving as my advisor and chairman of my dissertation committee. This research would not have been possible without his patience and constant guidance throughout the entire process. This debt of gratitude also extends to the Office of Academic Affairs who at a very crucial time allowed this research to continue.

The author is also very deeply appreciative for the unwavering support of O.S.U.'s Kerr Library faculty and staff who give new meaning to the role of "service" within an academic community. Special thanks goes to Marjorie Knittel-Educational Specialist par excellence. Her library orientation seminars, acquisition of professional journals in education, curriculum expertise, and eagerness/availability to share her vast storehouse of knowledge with others, has made quality research possible for students in education and a challenge for them to constantly improve. To Robert Lawrence for the many long and tedious hours involved with BRS data collection. To Don Unger whose multifaceted professional expertise and interests I have found a kindred spirit; and to Doris Tilles and members of the Inter Library Loan Department for obtaining a multitude of, and at times difficult to acquire, research documents.

TABLE OF CONTENTS

Table of Contents (continued)

LIST OF FIGURES

LIST OF TABLES

List of Tables (Continued)

List of Tables (Continued)

AN INVESTIGATION OF THE RELATIONSHIPS BETWEEN STUDENT EVALUATIONS AND FACULTY, CLASS, AND STUDENT DEMOGRAPHIC VARIABLES IN RATING INSTRUCTIONAL EFFECTIVENESS

CHAPTER I

INTRODUCTION

Bobbitt (1985), Chabotar (1989), Dennis (1986), Slaughter (1988), and Slaughter and Silva (1985) argued that most systems of higher education have experienced dramatic decreases in state and federal funds, and will face new periods of retrenchment in the 1990s due to diminishing financial resources. This is especially evident in the State of Oregon with the recent passage of Measure Five.

One consequence of this retrenchment has been a decline in tenure and promotional opportunities emanating from diminishing enrollments and uncertain budgets. This in turn has made faculty increasingly sensitive to the criteria governing their career advancement (Marsh, Overall & Kesler, 1979). Student evaluation results influence tenure and promotion decisions, salary recommendations and, when combined with a publication history and a service component, form the bases of advancement criteria for faculty in most institutions of higher learning (Machlup, 1979; and Marsh, Overall, & Kesler 1979).

Student evaluations of faculty teaching effectiveness are among the most commonly used and easily obtainable forms of data for advancing faculty in their careers; yet, they still remain one the most controversial components of the summative evaluation process. It has been suggested that the grades students expect to receive and/or such demographic variables as the gender of the faculty, class size etc., may actually influence faculty evaluation results (Aleamoni & Hexner, 1980; Doyle & Whitely, 1974; Marsh, Overall, & Kesler 1979; Tollefson, 1983).

In an interview on the Oregon Public Broadcasting System (OPBS), Chancellor Bartlett from the Office of the Oregon State Board of Higher Education warned, "The quality of instruction offered at Oregon's eight institutions is about to suffer"; the public was cautioned that faculty and academic programs are being substantially reduced but not the number of students at these institutions. Therefore, the ratio of students to faculty in lecture and laboratory sections is expected to increase with, what the Chancellor believes will be a corresponding decrease in the quality of instruction offered in these sections. Class size and quality of instruction may become a serious issue in Oregon higher education (Oregon Public Radio interview, April 9, 1991).

There is often suspicion that other variables unrelated to the quality of instruction also affect student evaluations. The extreme views of some critics concerning student evaluations have suggested that instructors need only give high grades and demand minimal student output to receive high ratings (Marsh, Overall, & Kesler, 1979).

These critics, especially faculty who are in the instructional trenches, further argue that variance in course ratings is associated with extraneous instructional variables such as course level and class size as well as extraneous student variables such as expected grade and reasons for enrolling in the course (Aleamoni & Hexner, 1980). Research on instructional processes where extraneous variables were treated as independent variables has shown varied results.

Tollefson (1980) argued that educational level, interest in the subject matter, and reason for taking the course have been shown to predict course and instructor ratings. Studies on rating of teachers by the students at different educational levels have yielded inconsistent results. For example, Aleamoni and Hexner (1980) cited eight studies which found no significant differences in instructor ratings given by students at different educational levels and eighteen studies which found graduate students and upper-division students giving higher instructor ratings than lower-division students. Dolyle and Whitely (1974) also reported conflicting results in studies assessing differences in instructor ratings given by lower- and upper-division students. For example, Costin et al. (1971) concluded that ratings given instructors by male and female students did not differ significantly. This was also reiterated in the findings of Basow and Distenfeld (1985), Bennett (1982), Elmore and LaPointe (1974, 1975)

However, Kaschak (1978, 1981) found a consistent bias in favor of male professors. Lombardo and Tocci (1979) reported that male professors were considered by male students to be more competent as instructors. Female students showed no sex bias. Harris (1976) reported that students preferred the instructor who possessed traits stereotyped as appropriate for their own gender and discipline.

Despite these inconsistencies in research findings, an increasing number of colleges and universities across the nation require results from student evaluation of faculty teaching performance to be included as an integral component of the summative evaluation process. In part, the impetus comes from students interested in shaping institutional policy. Although, the major force for the inclusions of student ratings in summative evaluation processes comes from faculty and administration genuinely concerned with improving instruction and creating a fair and equitable system of rewards.

Feldman (1983) argued that an insufficient amount of research has been conducted to substantiate the influences on student ratings by such demographic variables as faculty tenure status, academic rank, research productivity, and gender effects. Kaschak (1981) contended that inequalities exist on how students rate male and female faculty and that these results have a direct influence on tenure and promotion rewards. Research is needed on a wider range of students and colleges within universities to address the issues involved. The current investigation was conducted in response to this call. In studies where non-representative samples have been used the results tended to be conflicting and interpretations confused as to the effects of select demographic variables on faculty rating results. The problem is further compounded by the enormous variation in methodologies employed to accommodate non-representative samples from one study to the next and the type of statistic used to analyze these data once collected.

Statement of the Problem

In view of the growing use and importance of student evaluations, and the need for accurate data analysis to deal with the enormous changes that are taking place in higher education the present investigation was designed to answer the following questions: Are faculty ratings related to the faculty demographic variables of tenure status, research productivity (publications in refereed journals), academic rank, and gender and to the student/class demographic variables of expected grade, course major, class status, grade point average, percent of class attendance, gender, and reason for enrolling in a course?

Purpose of the Study

The purpose of this study was to analyze and interpret the results of Oregon State University's fall term student ratings of faculty teaching performance (n=40,000) in order to accurately determine whether relationships exist between a wide range of demographic variables and the ratings faculty receive from students enrolled in their classes. The specific goal of this investigation sought to identify potential relationships between a set of demographic characteristics or 12 independent variables such as gender of the student/instructor, major/non-major courses, required/elective courses, graduate/undergraduate courses, tenured/nontenure faculty, full-time faculty/teaching assistants; student expected grade, overall GPA, and percent of classroom attendance. The dependent variables (12 subscales, including the overall subscale) constitutes the Student Assessment of Teaching Instrument (SATI).

Significance of the Study

In times of retrenchment, identifying productive faculty and programs for the reallocation of funds becomes as important as the acquisition of new funding sources. Administrators are currently faced with the challenge of retaining instructional faculty who meet the tripartite role of research/publication, instruction, and service within the academic and Corvallis communities.

Public demands are strong for greater accountability of those involved in university administration, and for quality instruction, faculty, and a marketable student product. The institutional needs are equally strong for knowledge of demographic variable relationships and for the accuracy of SATI in gathering student rating data. In light of these needs the results of this investigation were used in the following areas:

1. to aid university teaching personnel involved with formative evaluation processes to improve instruction;

2. to make recommendations to central administration on how best to use faculty ratings for tenure and promotion decisions;

3. to provide data for the central administration to re-examine the tripartite role of O.S.U. faculty;

4. to aid administration in allocating resources for research productivity to faculty members who frequently publish and teach well; and

5. to establish a publication benchmark for college personnel directly involved with recruiting tenured and nontenure faculty.

In summary, collecting student evaluations of faculty performance across an entire university or college is a difficult task. Although, when the efforts of central administration makes it a requirement that **all** faculty with teaching responsibilities be evaluated by students it then becomes possible. Such is the case with the present study and one that has not been previously accomplished in any similar study.

Additionally, the present study will cover a broader range of faculty, student body, and departments than in all previous studies. This diversity may lead to different findings but more importantly, provide in-formation that researchers may be able to answer with greater confidence the question of whether select demographic variables are significantly related to teaching effectiveness of university instructors as measured by student evaluations.

Hypotheses

In order to determine the extent to which selected instructor demographic variables were related to student ratings, 13 null hypotheses were posited. The level of significance for each statistical test was .05 or less. The null hypotheses follow:

HO:1 There is no significant difference in course evaluations between faculty across major and non-major courses.

HO:2 There is no significant difference in course evaluations between faculty across required versus elective courses.

HO:3 There is no significant relationship between mean faculty ratings and student GPA.

HO:4 There is no significant relationship between mean faculty ratings and student expected grade.

HO:5 There is no significant relationship between mean faculty ratings and student percent of classroom attendance.

HO:6 There is no significant difference in course evaluations between faculty across graduate and undergraduate courses.

HO:7 There is no significant difference in course evaluations between male and female students.

HO:8 There is no significant difference in course evaluations between male and female faculty.

HO:9 There is no significant difference in student evaluations of faculty based on class size.

HO:10 There is no significant difference in course evaluations between tenured and nontenure faculty.

HO:11 There is no significant difference in the course evaluations of non-international graduate teaching assistants and full-time faculty members.

HO:12 There is no significant difference in course evaluations between non-international and international graduated teaching assistants.

HO:13 There is no significant relationship between course evaluations and publication rates among full-time faculty.

Limitations

This study will be conducted under the following limitations:

1. This study is limited to the degree that the Student Assessment of Teaching form is a valid and reliable measure of effective teaching.
2. This study is limited to the degree that students will answer with honesty questions on the Assessment of Teaching Instrument.
3. This study is limited to universities with similar demographic composition. Similar from the standpoint that OSU is an agricultural and technically oriented university; places a major emphasis on disciplines of science and a low emphasis on liberal arts disciplines; is a mid-sized university of 15,000 students, and is designated as a land-grant and sea-grant university.
4. This study is limited to those variables measured by the Student Assessment of Teaching Instrument which does not take into account such variables as instructor motivation, workload, and professional experience.

Definition of Terms

The following definitions are relevant to this study:

Scholarly Productivity: The number of articles faculty have published in refereed journals over a period of four years beginning with the 1985-1986 academic year and ending with the 1988-1989 academic year.

Collinear Dimensions of Effective Teaching: Those variables or dimensions of effective teaching that are similar in nature. In assessment of teaching instruments each subscale would be measuring essentially the same aspect of teaching if they were collinear, leading to redundancy and ambiguity in the questions being asked and confusion in the responses provided.

Refereed Journal: The American Library Association Glossary of Library and Information Science (TALAGLIS) defines a refereed journal as a serial or periodical evaluated by one subject specialist in addition to an editor before being accepted for publication (1983, p. 188). However, articles sent to the most prestigious professional journals have two or more "blind" reviewers in addition to the chief editor. The results of the reviews either allow the article(s) to be published, sent back with a note from the editor with suggested changes necessary for resubmission and/or rejected without the possibility of re-submission.

Student Ratings: The student perceptions of teacher effectiveness as reflected in their responses on the Student Assessment Instrument (SATI).

Tenured faculty: A faculty member who has been granted tenure status under existing university tenure policy (indefinite).

Non-Tenured Faculty: Faculty not formally granted tenure status under university tenure policy (fixed term).

Annual Tenure: Faculty on a tenure track but who have not been rewarded tenure status.

Teaching Assistants: Graduate students in degree programs employed by the university for teaching assignments, including a wide range of additional duties that may or may not be germane to instructional responsibilities.

CHAPTER II

ANALYSIS OF RELEVANT LITERATURE AND RESEARCH

Introduction

The number one goal of the 1989 Strategic Planning Process currently underway at Oregon State University for the 1990-1991 biennium is to "serve people through instruction, research and extension." Under this goal objective 1.1 of the revised draft states: "Strengthen programs and rewards that promotes and recognizes good teaching" (p. 7).

Additionally, the 10 institutional goals and the 65 subsumed objectives constituting the revised plan are integrated within four general themes listed by their order of importance pursuant to the stated goals and objectives. These themes include teaching, facilities, cultural diversity and university relations. Quoting from the teaching theme:

> Again and again, groups studying particular goals would independently arrive at the same conclusion: the recognition accorded excellence in teaching, and the rewards accorded to excellent teachers, must be substantially increased. This was found to be central to the key goals of serving people, aiding students in realizing their full potential, recruiting and retaining faculty committed to excellence, and recruiting outstanding students. To strengthen programs and rewards that promote and recognize good teaching is the top priority objective for the number one goal. (p. 4)

In 1989 the Committee for the Advancement of Teaching (CFAT) was appointed to review the summative evaluation process at Oregon State University for the purpose of creating a systematic and fair faculty evaluation process on which to base personnel promotion and tenure decisions. Dissatisfaction with the enormous variability in faculty evaluation strategies across campus, prompted CFAT to look for a standardized system that recognized effective teaching in this evaluation process. The goal is to achieve a balanced evaluation within the tripartite role of the faculty, e.g. research, instruction and extension in the service to the state.

The President of Oregon State University, Vice President of Academic Affairs and, chairperson of CFAT are all committed to the belief that effective teaching can be rewarded only if it is first acknowledged as an integral component in the overall evaluation process and, secondly, if it is elevated to the status accorded research and extension at the university. The challenge to CFAT and the scope of this dissertation is to make recommendations to the central administration at Oregon State University based on how best to use the results of student evaluations in establishing new faculty evaluation guidelines.

Peter Seldin (1984) opens his book Changing Practices in Faculty Evaluations with the following observation:

> Faced by an economic squeeze unprecedented in recent years for its severity and duration, the nation's colleges and universities are struggling to cope with reduced budgets, hunting for new money sources and casting a gimlet eye on which faculty to promote and which courses to teach. (p. 1)

Seldin (1984) argued that during periods of retrenchment universities react by focusing greater attention on mission statements, goals and objectives by allocating and reallocating scare resources and by examining the guidelines affecting promotion, tenure, retention, and merit pay increases at the institutional level. Boland and Sims (1988) believed that today's academic realities have provided the impetus for administrators and faculty to critically examine evaluation processes at their institutions. Importantly, evaluation serves as a means of documenting levels of productivity and quality of faculty performance.

The public interest in fiscal accountability among the institutions of higher education has resulted in a call for a comprehensive evaluation of the professorial ranks at these same institutions. Tax payers want to be reassured that their money is being well spent on academic programs contributing to the revitalization and continued growth of their state. They also demand academic quality with positive outcomes and an effective way of evaluating faculty performance to insure accountability (Chabotar, 1989; Chadwick & Ward, 1987).

Murray (1984) summarized the root causes of the growing sentiment for wise management of fiscal and human resources in the following observation:

> Evaluation of the teaching competence of individual faculty members has taken on increased importance in North American universities over the past 20 years. The reasons for this trend include, among other things, demands for accountability in public institutions, legal precedents requiring "due process" in evaluation of faculty performance, and the refusal of today's consumer-oriented student to tolerate substandard teaching. (p. 34)

The notion of legal documentation of faculty performance and institutional resource utilization acceptability becomes even more crucial when the courts are involved. Gillmore (1983-1984) noted a trend in "the 1970s and 1980s boom in litigation by faculty challenging negative employment decisions." Hendrickson and Lee (1983) stated that:

> An uninhibited evaluation of a colleague's professional competence became increasingly hazardous within colleges and universities during the 1970's. For several reasons, a decision to deny any educator promotion or tenure or to terminate his or her employment is now more likely to result in legal action than in previous years. (p. 10)

In recent years the courts have displayed an increased willingness to intervene especially in cases where careful documentation and the application of uniform procedures have been lacking in faculty promotion and tenure decisions. Gillmore (1983-1984) noted that thus far the courts have tended not to question criteria for promotion and tenure; "Generally speaking, courts are more likely to review the fairness or reasonableness of the application of the decisional criteria rather than evaluating the relevance or appropriateness of the criteria themselves" (p.561).

Olswang and Fantel (1980-1981) contended that periodic performance review is not incongruent with either tenure or academic freedom. This practice, they claim, "would also result in a sensible and constructive alternative to the other and more drastic measures institutions have at their disposal to resolve problems of faculty decline and institutional inflexibility" (p. 4).

Individual components of the evaluation process such as peer reviews, publication rates, research productivity, student ratings, self-rating, and curricula content are all viewed as valid and objective criteria for judicial scrutiny when litigation is brought against an institution by a faculty member. Aleamoni (1986) noted that most instructional evaluation systems consist primarily of student evaluations while Gillmore (1983-1984) observed that if student ratings do in fact qualify as evidence in support of faculty employment decisions, the systematic analysis and interpretation of their results are of extreme importance, especially when the reliability and validity of student rating instruments become an issue.

Lee (1985) summarizes the responsibility of academe, with full support of judicial precedent, in determining what criteria are appropriate to apply in evaluating faculty performance, and the quality of that performance. She comments:

> It would benefit higher education and justify such judicial confidence in academic autonomy, however, if higher education strove more vigorously to link the requirements for advancement in academe--the qualification for promotion and tenure--to improvement in the quality of the education being provided the students and the services being provided to the university, the profession, and the public at large. (p. 364)

Ormrod (1986) argues that at many institutions once faculty members attain tenure status and or promoted to higher levels they are not formally evaluated again. In traditional systems of tenure and promotion, an inherent shortcoming is a decline in work effort and research productivity of tenured "Professor" status. Bennett and Chater (1984) comment, "It is the performance of the individual who is tenured that is under evaluation, not the tenure of that individual" (p. 39). An intuitive assumption that permeates the literature is that the older faculty members become the less effective they are as teachers.

Central to the issue of accountability and fulfillment of the tripartite role of instruction, research and service is the level of faculty productivity, which Centra (1983) defined as the quality and quantity of research and publication. Astin and Lee (1967) and Centra (1977) note that although teaching, research, and service are the major functions of most universities, performance as a researcher or teacher generally receives the most attention in determining rewards and promotions. Holbert (1985) argues since higher education is a labor-intensive industry, one way to decrease costs is by increasing the productivity of the faculty.

Centra (1983) argues that most faculty members believe that research benefits teaching, especially at the graduate level where contact hours with graduate students conducting their own research tends to be quite high. This dictates that major professors be competent in content knowledge and the fundamentals of conducting university level research and that they demonstrate the ability to publish results in major refereed journals in their respective disciplines.

Ladd (1979) extends this contention further by arguing that research, meaning "scholarly activity aimed at the creation of new ideas, new know-ledge, new art forms" (p. 8) is the primary mode of scholarly exchange, which generally takes the form of publication of articles, books, and monographs. Jencks and Riesman (1969) provide the following rationale:

> Publication is the only way a man [sic] can communicate with a significant number of colleagues or other adults. Those who do not publish usually feel they have not learned anything worth communicating to adults. This means that they have not learned much worth communicating to the young either. There are, of course, exceptions: men who keep learning but cannot bring themselves to write ... Still these are exceptions. (p. 532)

Holbert (1985) noted that it was no coincidence that Jencks and Riesman used the male gender when describing college faculty. It is predominately men that most of the research on faculty productivity has been conducted. Only recently has the focus shifted to include women faculty.

If publishing is an important dimension, or requirement, of academic life, then it seems imperative that the productivity of both males and females be considered. Blau (1973) argues that the reputation of an academic institution is strongly correlated with the research productivity of its faculty. All officials of higher education are concerned with the reputation of their institutions in both public and political arenas. In periods of budgetary restrictions and the severe competition for dwindling dollars that follows from it, forces institutions to make an all-out effort to appear highly productive in all spheres of influence.

Centra (1979b) notes a "spillover" effect in enthusiasm and knowledge acquired from conducting research and publishing. Dispelling misconceptions and providing insights into stubborn problems create the rich seedbed of ideas on which the next generation of scholars flourish. Glueck and Jauch (1975) surveyed natural science faculty and department heads at one university. The results indicated that 95% of the faculty agreed with their findings, they commented, "that research increases teaching effectiveness by increasing awareness and currency" (p. 29). Currency refers to the process of remaining current, or up to date in a discipline via research and publication.

Findings from the few investigations conducted to uncover the relationship between publication rate and teaching effectiveness, revealed only conflicting results. Voeks (1962) found no statistically significant correlation between teaching effectiveness as determined by student ratings and membership in a university research society or with publication rates for a wide range of academic ranks and disciplines. While the data suggest that publishing and effective teaching do not go hand in hand, the data did not show that publishing was done at the expense of effective teaching, or vice versa.

An often heard intuitive assumption is that the frequency of research and publication is related to the level of poor teaching performance. In the few cases where this appears to be true, the cause may be inadequate training and organization in classroom management. Content expertise does not automatically confer teaching competence (Good & Brophy, 1984).

Investigations where faculty opinion on teaching performance and publication rates were compared produced similar results. Aleamoni and Yimer (1973), Hayes (1971), and Hoyt (1974) used a weighted measure summary of books and articles produced to estimate research productivity and reported no correlation between research output and peer nominations of good teachers.

Performance evaluation has historically been a part of an academic career, particularly of non-tenured faculty, and currently with increasing frequency as the bases for making tenure and promotion decisions effecting faculty with tenure. Bennett (1985) argues that there are enormous mutually-shared benefits in developing and implementing a systematic post-tenured evaluation for both faculty and institutions alike. Institutions achieve the kind of "academic excellence" often spoken about and maintain public perception of integrity and constituency support, while generating and protecting the most favorable conditions for teaching and scholarship.

Chait and Ford (1982) argue that periodic evaluation creates opportunities for regular reflection on the adequacies and inadequacies of the teaching-learning process, as well as on appropriate teaching strategies. Areas of pedagogical weakness and strengths can thus be identified and examined systematically in effective teaching. It also plays an important role in departmental development and establishing future directions. Chait observed:

> The evaluation process subtly, almost subconsciously, creates an expectation of progress and advancement. Properly executed, the process also enables individuals and departments to set directions and priorities in harmony with institutional objectives. (p. 34)

The results of the few studies available on differences in student perception and rating of male and female instructors suggest that gender-stereotypic qualities play a significant role in the way students rate college teachers. Basow and Silberg (1987) caution the reader in the interpretation of the results of these studies. They state:

> Their findings are unclear because in some cases professor sex typing interacted with student sex or professor sex. Furthermore, studies have varied considerably in methodology, with some using written descriptions and some using videotaped or real life classroom instructors...given the confusion in the literature, further examinations of the effects of sex and sex typing on student ratings of college professors is warranted. (p. 308)

Additional evidence in support of further investigation arises from the difficulty with acquiring a representative sample of an entire faculty. The studies by Basow and Howe (1979) and Basow and Silberg (1987) are perhaps the most extensive studies on gender differences. However, the first study surveyed only a small fraction of tenured professors in a college setting and totally excluded those in engineering. The faculty sample size was too small (n=16) for an adequate comparison across departments, and while the student sample size (n=1000) was relatively large it represented only slightly more than half of the student body.

The 1987 Basow and Silberg study attempted to replicate the 1979 Basow and Howe study. The outcomes were similar because the problem of securing a representative sample could have significantly influenced the results in each study.

An important problem related to effective teaching in higher education is the quality of education provided at the undergraduate level. Concerns about this quality have been voiced in national curriculum reports published in the Chronicle of Higher Education (1985). These reports attacked administrators and faculty for the "steady" decline in the quality of instruction available to undergraduates. Currently, it is a major issue, especially as it relates to foreign graduate teaching assistants and their ability to meet academic demands required in university-level instruction (Constantinides, 1987; Heller, 1985c).

The linguistic, pedagogical, and cross-cultural difficulties encountered by American students from foreign TAs are collectively labeled the "foreign TA problem" (Bailey, 1982; Constantinides, 1987). The problem appears pandemic in U.S. institution of higher education where foreign teaching assistants are employed to teach undergraduate courses.

Their performance has become an emotional issue generated by angry parents and students in the form of complaints and demands for action from state boards of higher education and state legislatures across the nation. Place these political concerns in the academic arena along with student concerns (and those of their parents) about the steady increase in the costs of acquiring an education and a confrontation emerges (Alciatore & Eckert, 1972; Bailey, 1977, 1982,; Barrus, Armstrong, Renfrew, & Garrard, 1974; Bray & Howard, 1980; Brinton & Gaskill, 1979;

Buckenmyer, 1972; Carroll, 1980; Damarin & West, 1979; Staton-Spicer & Nyguist, 1979).

Oregon was not exempt from the deluge of citizenry discontent. The Chancellor's Office through the State Board of Education was requested by the State Legislature to investigate the seriousness of the problem and act accordingly. The Strategic Plan of Oregon State System of Higher Education 1987-1993 made the following recommendation:

> Recommendation 9: Institutions will establish guidelines and evaluation procedures for the selection and retention of graduate teaching assistants that assure they meet high teaching standards. (p. 38)

Bennett (1984) states in To Reclaim a Legacy: Text of Report on Humanities in Education:

> In providing better quality, faculty and instructors should perform at the highest level of competence. Students are entitled to receive an education which links qualitative growth with effectiveness. (p. 19)

Teaching assistants often teach the largest portion of introductory courses. Davis (1984) reported that in some universities, like the University of Wisconsin-Madison, all beginning and intermediate language courses were taught by TAs. Franck and Samaniego (1981) estimated that over 90% of the introductory courses at the University of Michigan were taught by TAs. Ohio State University estimated 25 percent (493) of its 1,974 TA compliment were foreign students. A conservative dollar investment for employing 1,974 TAs was placed at close to 20 million dollars per year with the inclusion of tuition waivers in the calculation. The value of quality instruction for the money spent is not only the right of the undergraduate but also the employing institution as well.

Fall Quarter of 1989 at Oregon State University in the College of Science the percentage of foreign TAs was approximately 38. Eighty percent of TA's were from Asian countries (Office of the Vice President for Personnel Services, Oregon State University, 1989).

Staton-Spicer and Nyguist (1979) argue that political and financial investments are not the only reasons for the improvement of teaching. Although, on these grounds alone, the authors assert, improvement of undergraduate instruction needs no justification. Prudent utilization and management of human resources, like financial resources, will increase the competitive edge of institutions in their efforts to recruit and retain top students and faculty. Armenti and Wheeler (1978) place effective teaching as the critical component in competition among institutions for undergraduate students.

The degree of preparation TAs make will have an impact on institutions of higher education in ways other than finances. The TAs are responsible for a significant portion of the instruction received by undergraduates, and they usually teach courses which are often the undergraduates' first contact with a particular department. Thus, TAs should reflect the nuances of a discipline and do so effectively if retention efforts prove to be effective.

First impressions often endure over long periods of time in spite of a preponderance of evidence to the contrary. Barlow (1985) showed that first impressions made by beginning biology students about their TAs did not significantly change over a semester even thought TAs made significant changes in their teaching approach to enhance student learning.

Faculty evaluation processes are central to promotion, tenure, and merit pay issues, and they should be credible when making critical decisions affecting the future of faculty. Bobbitt (1985) argues that any systematic plan of evaluation should have the concomitant of both faculty and administration if it's to succeed. The plan is essentially a shared belief in its relevance and mission to provide tangible rewards. Importantly, the process should include a variety of data from all possible sources reflecting faculty performance in areas identified and agreed upon by both the institution and school as being substantially valid indicators of performance.

Bland and Schmitz (1988) in an extensive survey of the research literature involving faculty or institutional renewal programs generated 152 specific recommendations from over 135 authors in 288 referenced sources. The most frequent recommendation mentioned, "Stressed the essential, critical link between faculty development and institutional mission and policies, a link that can only be achieved if each party acknowledges its role" (p. 196).

Researchers in the area of faculty evaluation and development such as Aleamoni (1987), Arnst (1978), Bevan (1985), Buhl (1975), Centra (1983), Chait (1979), Eble and McKeachie (1985), Lindquist (1978), Peterson (1981), Rice (1979), and Seldin (1984), have posited the essential integration of faculty and institutional goals. The following statement of Chait and Ford (1982) serves as one example:

> Determine what faculty and staff as individuals seek and value....Make certain that the performance to be rewarded supports institutional goals; insure that institutional rewards reinforce individual desires. None of the links between goals, performance, rewards, and desires can be broken for the system to be effective. (p. 231)

Buhl (1982) also makes the following point:

> Professional development [and evaluation] should be seen as a system through which institutional and department goals and personal and professional goals are matched so that the former may be achieved by facilitating the later. (p. 65, brackets mine)

Thomas (1980) blends the common goals of faculty development and faculty evaluation. When the elements of both are integrated they become mutually supporting, both the individual and the institution stand to benefit, in terms of improved individual organizational performance. Hammons (1983) notes that when both faculty evaluation and development programs are absent, the results are often better than when only one of the two is present. Hammons argues further that faculty evaluation systems without a faculty development program is "doomed" to eventual failure.

Centra (1979a), Schwab (1975), Gaff (1976), and Seldin (1980,1984) assert that universities experience short-term gains when either faculty evaluation or development programs are implemented in isolation. A summative evaluation requires the means of correcting instructional problem areas once identified. Conversely, a formative evaluation process requires the means of rewarding faculty efforts to improve instruction. Arreola (1979) makes a simple yet poignant observation about the issue of the validity of faculty evaluation systems: "If the faculty evaluation system or program...is considered by the faculty [including administration] to be fair and useful, it is functionally valid regardless of its validity in some statistical or psychometric sense" (p. 241).

Periodic evaluation of faculty (especially those who are tenured) is not without its opponents. The American Association of University Professor (AAUP) states without qualification and/or defense in the November/December 1983 issue of <u>Academe</u> that periodic evaluation would bring scant benefit, would incur unacceptable costs, not only in money and time but also in a damp-ening of creativity and of collegial relationships, and would threaten academic freedom. The position statement concludes with the following remark: "[t]he results of any system of evaluation should not be allowed to be used as grounds for dismissal..." (p. 12).

This position is a far cry from the AAUP position statement of 1915 edited by John Dewey:

> If this profession [higher education] should prove itself unwilling to purge its ranks of the incompetent and unworthy, or to prevent the freedom which it claims in the name of

> science from being used as a shelter for inefficiency, for superficiality, or for uncritical and intemperate partnership, it is certain that the task will be performed by others-- who lack certain essential qualifications for performing it, and whose action is sure to breed suspicion and recurrent controversies deeply injurious to the internal order and the public standing of universities. (p. 132)

Bennett (1984) contends that the issue is not whether there will be a comprehensive evaluation structure established at a particular institution for tenured and nontenure faculty, but when and by whom. The issue is whether the evaluation process will be created and periodically reviewed by faculty and administration within an institution, those best qualified to construct appropriate procedures, or alternatively, imposed by external agencies like boards of regents, boards of education, legislators, and accreditation committees. Bennett (1985) argues that systematic and fair evaluation of faculty performance can preserve the integrity of academic freedom while assuaging public and institutional concerns about the vigor, productivity and accountability of faculty.

Summative and Formative Evaluation Analysis

McKeachie (1986) makes a crucial distinction between summative and formative strategies as it applies to instructional evaluation. In making personnel decisions, administrators need some index of overall teaching effectiveness in order to determine if a particular faculty member will be promoted or rewarded with a merit pay increase based on teaching excel-lence. Formative evaluation on the other hand, serves as a diagnostic tool designed to improve instruction through identification and correction of problem areas.

Few authors will argue with the notion that a comprehensive summative evaluation system should be multi-dimensional. In discussing effective teaching Centra et al. (1989) states that "evaluating teaching is an es-sential element in the promotion and tenure decision process and must be done in a manner that is fair, comprehensive, and cost effective" (p. 1). Seldin (1984) extends this criteria to all major components of the summative evaluation process.

Because of the plethora of literature on the subject--in access of 380 documents--the author has chosen to provide a general review of the most salient aspects contained in the literature base. Seldin (1980; 1984) contends that a balanced summative evaluation would contain the following components: self-evaluation, peer evaluation, student ratings, evaluation of curriculum material, review of classroom records and activities by faculty and administrators, and alumni surveys.

Self-evaluation serves as a reference point to compare results with other components of a summative/formative evaluation process. It is also a logical first step toward pedagogical improvement to identify the strengths and weaknesses of instructors. The greater the gap between self-evaluation (perceptions) results and the data from other sources, the greater the impetus to make appropriate changes (Centra, 1973c; Blackburn & Clark, 1975).

Student-rating data when compared with self-evaluation data provide solid evidence on which comparisons can be made, especially when professors use the same student rating forms to assess their own teaching. This method is cost effective and requires approximately 20 minutes per quarter to accomplish.

However, self-evaluation is of minor importance in the enhancement of instructional effectiveness, if the professor is not given the actual student assessment of teaching data set to review and reflect upon. It is important at this juncture to also provide "expert" consultation concurrently with an interpretation of the data.

Suggestions should be made on how to transform student evaluation results into teaching behaviors that will increase classroom teaching effectiveness. The body of research findings on this subject suggests little if any instructional improvement can be expected when there is an absence of additional help in correlating self-evaluation data with student evaluative data (Cohen, 1981; Stevens & Aleamoni, 1985).

Aleamoni (1976) and Piper (1983) caution readers that faculty may be fearful of self-incrimination in the presence of unsympathetic administrators and colleagues, and therefore they tend to over exaggerate their actual teaching ability. Probably for this reason, self-evaluation is not highly regarded in making promotion and tenure decisions because the risks are too high.

Regardless of this apparent draw-back, Aleamoni (1986) strongly believes in self-evaluation as an integral component in formative evaluation processes, but considers it less effective in summative evaluations. He contends that an emphasis should be placed on the potential to promote reflection on teaching as well and to bring into concomitance those perceptions held by students and faculty on actual teaching effectiveness. Further, it should not be the primary source in which promotion and tenure decisions are based but, rather, utilized for the purpose of corroborating and/or substantiating findings from other components of the summative evaluation process.

Peer evaluation processes generally involve faculty and administration. They make routine classroom visitations for the purpose of observing faculty classroom teaching activities and examining instructional materials used by an instructor during a particular phase of instruction.

Seldin (1984) and Moffatt (1989) recommend senior colleagues be chosen by the department chair to conduct evaluations of faculty up for review, or by an academic dean in consultation with the department chair. From a list of six colleagues chosen by the instructor up for review, three would be selected by the department chair to form the evaluation committee. It was also recommended that several days prior to the visitation, observers should meet with the instructor to review the content of instructional material and teaching methods/strategies planned for the class to be observed.

Individuals and committees assigned the task of evaluating the teaching skills and effectiveness of faculty should be knowledgeable of research on the practice of teaching. Content expertise and years of teaching experience alone do not automatically confer the ability to teach effectively in a wide range of situations (AAHE, 1984; Baldwin, 1983; Becker, 1981; Bergquist & Phillips, 1975; Centra, 1978; Eble & McKeachie, 1985; Gaff, 1976; Hoffman, 1975; Lindquist, 1978).

Studies have revealed a positive relationship exists between colleague ratings of instructional effectiveness and student ratings. Aleamoni and Yimer (1973) reported correlation coefficients at 0.70. Investigations by Stallings & Spencer (1967) noted the same magnitude, 0.70. These same correlation coefficients were also reported in replicated studies by Linsky and Straus (1975) and Marsh (1984).

A distinct limitation of peer evaluation is that it may not be cost-effective in terms of additional time required from an already burdened faculty. A number of researchers, such as Aleamoni (1987), Centra (1975; 1976; 1979a), Chait and Ford (1982), and Seldin (1984), recommends multiple classroom observations, both announced and unannounced, before an accurate assessment of effective teaching can be made. They also suggest the number of visitations should range from 3 per semester to no fewer than 10 per academic year.

Findings from several investigations indicated that when given an option, faculty preferred student evaluation results over those of colleagues (Centra, 1975; Grasha, 1977; Piper 1983). This sometimes illustrated a lack of confidence in a colleagues' ability to evaluate teaching and reflected the fact that student evaluations are often more positive [see Student Ratings Reliability and Validity, below]. The fact that correlation coefficients are similar between colleague and student rating data does not suggest that faculty ratings are any less accurate, only that student rating data in summative evaluation processes were preferred over data generated by colleagues in the above studies.

Currently there are approximately 328 publications which, to varying degrees, address the issue of student ratings. When properly developed and implemented student ratings provide the most reliable and cost-effective means of obtaining evaluation information (Stevens 1985; 1987).

There is substantial evidence that student ratings have a high degree of reliability. Guthrie and Costin (1954) and Costin et al. (1971) conclude that student judgments tend to be relatively stable, and Cooper and Petrosky (1976) found that even secondary students tend to be consistent in what they say about instructors and instruction. Abrami, Leventhal, and Perry (1982) investigating the influence of instructor personality on student ratings found that students were adept at discriminating between various aspects of teaching ability and the "nice, friendly, humorous" atmosphere generated by the instructor in the classroom.

Studies, where alumni from Purdue University were asked to rate their former instructors five to ten years after graduation, reported correlations of .80 and .90 between the assessments taken in class and those years later (Drucker & Remmers, 1950;1951). Marsh (1977) at the UCLA and Aleamoni and Yimer (1974) at University of Illinois replicated Drucker and Remmers study using graduating seniors. They reported similar results. These studies suggest that students do make consistent judgments of instructional effectiveness while in the class and that these judgments appear to endure years later. Barlow (1985) noted that over time first impressions were the most enduring.

Cohen and Berger (1970), Frey (1975), Frey, Leonard, and Beatty (1973), and White, Hsu, and Means (1978) reported a fairly high positive relationship between objective measures of learning and the way students rate teaching performance of their professors. Student ratings of faculty teaching performance can lead to improved instruction (Aleamoni, 1978; Aleamoni & Hexner, 1980; Centra, 1979a, 1980; Cohen, 1980b; Eble & McKeachie, 1985; Kulik & Kulik, 1974; Marsh, 1980a; McKeachie, 1979; Millman, 1981; Overall & Marsh, 1979).

Stevens and Aleamoni (1985) demonstrated that feedback from a consultant in conjunction with standard computerized output data from student ratings significantly improved classroom teaching effectiveness. Computer application evaluation processes greatly accelerates the location and identification of problem areas in instruction as indicated by student rating results.

This student-rating data base creates a starting point from which consultants can make suggestions and initiate correction procedures. Stevens (1987) notes that not all instructors possess adequate skills and knowledge to understand, interpret, and apply feedback data either to improve instruction or to make summative evaluation recommendations. He recommended only those individuals with such ability be allowed to serve as consultants.

A number of studies provide evidence that course evaluations in mid-semester can bring about changes in teaching practices (Centra 1972; Gage 1974; McKeachie & Lin, 1975; Pambookian, 1976; and Whitman & Schwenk, 1982. Gill (1987) argues that positive feedback and evaluation process reduce anxiety and improve a professor's self-concept.

Aleamoni (1987) argues that most current instructional evaluation systems consist of only one component: student evaluations. He believes the low cost and expedience involved with collecting student rating data may result in an over reliance on this data base. This is especially true when time and money commitments are considered to acquire data from other components in the summative evaluation process.

Other important considerations in the summative evaluation process involve the enormous variability encountered in the implementation of university-wide student evaluation programs, which tends to reduce the reliability of the data set generated. Variables like student knowledge of whether their evaluations of instructor teaching performance will have an impact on decision-making processes, if the instructor is present or absent during the evaluation, the time evaluation is given in the semester, behavior of the instructor prior to an evaluation; all have a significant effect on evaluation results (Aleamoni & Graham, 1974; Aleamoni & Yimer, 1974; Centra et al., 1989; Cohen, 1981; Marsh, 1984).

Aleamoni (1987) concludes from analysis of the summative evaluation research data that administration has no better idea of the teaching effectiveness of their professorate than do students and faculty, yet most promotion and tenure decisions are made exclusively by administration. Such conclusions are substantiated by reports that correlation coefficients of .26 to .35 were generated when a comparison of teaching evaluations of instructors by administrators were compared to those of their colleagues and students.

Student Ratings Reliability and Validity Issues

Stevens (1987) argues that student-rating instruments, when properly developed, provide the most reliable and cost-effective means of obtaining evaluation information. Most institutions collect student evaluation data but the extent to which the data are used for summative evaluation purposes varies considerably from one institution to another. Aleamoni (1987) points out that most teaching evaluation systems currently in existence consists of only student evaluations.

Murray (1984) attributes the popularity of student ratings to a combination of factors, including political, expediency, and research support, as well as to the inherent difficulties found in alternative methods. But this has not always been the case. Bejar and Doyle (1978) estimated that the proportion of universities using formal student ratings as a measure of teaching effectiveness had increased from 35% in 1951 to 48% in 1968 and to 68% in 1975. Similarly, Seldin (1980) reported that in five years the percentage of 410 private liberal arts colleges using student evaluations on a regular basis in promotion and tenure decisions had increased from 29.1% in 1973 to 53.1% in 1978.

The 1978 survey broadened the inquiry to include both private and public liberal arts colleges. Seldin (1984) replicated the 1978 survey and noted an increase from 53.1% in 1978 to 66.6% in 1983. The use of student ratings as a means of evaluating course and instructor teaching effectiveness had increased substantially in ten years.

For personnel purposes, faculty and administrators have often voiced concern about the validity and reliability of evaluation data when used for such far reaching purposes as faculty promotion in rank, awarding tenure, determining merit pay increments, or as a starting point for instructional improvement.

There is a substantial body of research on the issue of validity of student rating data. Aleamoni (1987) lists eight general concerns voiced by faculty:

1. students cannot make consistent judgments concerning the instructor and instruction because of their immaturity, lack of experience, and capriciousness;
2. only colleagues with excellent publication records and experience are qualified to evaluate their peers' instruction. Reference was made to the contention held by W.E. Deming (1972) that these were the only ones that could qualify as good instructors;
3. most student-rating "schemes" are popularity contests with the warm, friendly, humorous, easy-grading instructor emerging as the winner every time;
4. students are not able to make accurate judgments concerning either instructor or instruction until they have been away from the course, and or institution for several years;

5. a general indictment of student-rating forms such that the data generated are both unreliable and invalid;
6. any of several extraneous variables could affect student ratings: class size, gender of student and instructor, time of day of course offering, whether the class is required or an elective, student is a major or non-major, term or semester that the course is offered, course level (freshman,sophomore, junior, senior, or graduate), and rank of the teacher from instructor to full professor;
7. the grades students either expect to or actually receive are highly related to their ratings of both the course and instructor; and
8. how can student ratings or evaluations possibly be used to improve instruction.

These "extraneous" variables subsumed under number six of Aleamoni's (1987) list, above, are in part, reiterated by Centra (1979) in that student, instructor, and course characteristics can influence rating:

1. discussion classes get higher ratings than lecture classes;
2. students tend to give higher ratings to courses that are major requirements or electives as compared to courses that fulfill a college requirement;
3. teachers are rated the poorest in the first year of teaching. There was a slight decline in ratings for teachers with over 20 years experience;
4. small classes of less than 12 students get especially high ratings and are also less reliable in their ratings (Centra & Creech, 1976);

5. no difference was found among faculty holding different academic ranks (Centra, 1976);

Variables Influencing Student Ratings

As mentioned earlier in this chapter, the results derived from student ratings are utilized to make comparative judgement of an instructor's instructional effectiveness for the purpose of improving instruction and for making summative evaluation decisions. Remmers (1963) notes that student perceptions recorded in teaching performance ratings may be influenced by factors other than an instructor's actual teaching ability. Therefore, to improve the validity of student ratings research, efforts should be directed toward the identification of those variables which might produce random errors in student ratings.

The value of student ratings depends upon taking systematic errors or bias into account in the design and analysis processes. This is accomplished through an understanding of how background variables, e.g. demographic, influence or infuse unsystematic errors into student responses on O.S.U.'s Student Assessment of Teaching Instrument, and/or how errors may be amplified in the interpretation of faculty rating results.

The majority of research on student ratings of instructors has been concerned with the effects of demographic variables on these ratings. Demographic variables in this study that might influence faculty rating results are: major versus minor courses, class size, required versus elective courses, gender of the student and instructor, expected grade, class attendance percentage, grade-point average, faculty rank, and student class level. These demographic variables were selected for investigation because they appear beyond the instructor's ability to control for them.

Anyone of these demographic variables has the potential of invalidating evaluation systems based on student ratings which do not adjust for their influence. A brief review of the literature on each variable mentioned above follows.

Major versus Non-major Courses

Aleamoni and Thomas (1977), Cohen and Humphreys (1960), and Rayder (1968), indicate that there are no significant relationships between student ratings and major and minor courses.

Required versus Elective Courses

Several investigators have found that students who are required to take a course tend to rate the instructors lower than students who elect to take courses (Cohen & Humphreys, 1960; Gillmore & Brandenburg, 1974; Pohlmann, 1975). This finding is supported by Gage (1974) and Lovell and Haner (1955) who found that instructors of elective courses were rated significantly higher than instructors of required courses.

In contrast, Heilman and Armentrout (1936) and Hildebrand and Armentrout (1936) reported no differences between the ratings of students in required courses and those in elective courses. Recent findings at Oregon State University (CFAT, 1989) indicate a positive correlation exists between the student ratings instructors receive within elective and required courses. In addition, these same data showed that there was no significant difference between class size and instructor ratings.

Instructor Academic Rank

Some researchers report that instructors of higher rank receive higher student ratings (Clark & Keller, 1954; Downie, 1952; Gage, 1974; Guthrie & Costin, 1954; Walker, 1969); however, others report no significant relationship between instructor rank and student ratings (Aleamoni & Graham, 1974; Aleamoni & Thomas, 1977; Aleamoni & Yimer, 1973; Linsky & Straus, 1975; Singhal, 1968). Conflicting results have also been found when comparing teaching experience to student ratings. Rayder (1968) reported a negative relationship, whereas Heilman and Armentrout (1936) found no significant relationship.

Gender of Students and Instructors

Another area showing conflicting results is in the existence of gender biases in the instructional evaluation of college and university. A number of studies report little if any significant differences in the evaluation of male and female professors on the basis of gender alone (Basow & Distenfeld, 1985; Bennett, 1982; Elmore & LaPointe, 1974,1975).

However, Kaschak (1978) investigated student evaluations of professors' teaching methods as a function of teacher gender, student gender, and academic field denoted as masculine, feminine, or neutral. Her results showed a consistent bias in favor of male professors. Lombardo and Tocci (1979) reported that male students perceived male instructors as more competent than female instructors, whereas female students showed no gender bias.

Kaschak (1981) noted that gender bias disappeared when professors were described as being award winning, although male professors were still described as being more "powerful and effective" than female professors. Additionally, female professors in "feminine" fields were rated as more concerned and likeable than male professors in those fields.

Harris (1975,1976) investigated the effect of instructor's gender, teaching style, and department on student evaluations of instructor performance. In these studies written profiles of male and female teachers were used, including adjectives described as stereotypically masculine or feminine. Both male and female college students rated instructors with "feminine" teaching styles more favorable on all variables except warmth.

In a 1976 study, Harris investigated the extent of masculine-stereotyped behavior in a broad range of learners. Students at all grade levels (nursery school through college) tended to evaluate more favorably masculine-stereotyped teachers on all variables except warmth and superiority. At the college level, professor gender-typing interacted with student gender, thus suggesting that students of either gender preferred the instructor who possessed traits stereotyped as appropriate for their own gender.

Foreign Graduate Teaching Assistants

Caramagno (1981) referred to the dual role of the teaching assistants (TA) as one of being an "academic hermaphrodite, a species sporting parts of both academic sexes--students and faculty" (p. 2). Gurnick (1981) describes the paradox of being a TA. The author concludes:

> A TA is a kind of middle person in the educational institution...a sort of apprentice instructor, but as one whose function is to enhance their [students] learning experience more so than to learn how to teach. (p. 3)

Paulston (1974) notes that in spite of the central role of a TA, relatively few are hired specifically for their abilities or interests in future teaching careers. Stockdale and Wochak (1974) also argue that many universities have used TAs as a means of "providing undergraduate instruction, and of providing financial support for graduate students, not as a means of training future college teachers" (p. 345). Lnenicka (1972) in a scathing review of the TA system claimed that using TAs as anything other than assistants is detrimental to the educational process:

> The undergraduate student and his [sic] parents, who suffer financial strain in order to provide for their children's college education, have a right to feel cheated and resentful when they

> find even one of the important courses in the undergraduate curriculum being taught by a graduate student, one who, in all probability, is inexperienced, unrehearsed, untrained for teaching, and who primary interest lies not in his [or her] teaching, but rather in satisfying the requirement for his [or her] degree. (p. 97)

Yet, many of these TAs on receipt of a Ph.D., will join the ranks of a professorate engaged primarily in research and teaching activities. Other academicians have shared Lnenicka's concern. Staton-Spicer and Nyquist (1979) reported an emerging trend of "growing concern about the improvement of the teaching effectiveness of graduate teaching assistants." (p. 199)

The TAs assist professors in grading exams and preparing course materials but, more often than not, are directly involved in undergraduate instruction (Bailey, 1982). Many are responsible for supervision of laboratory experiments, lead discussion sections which compliment professors' lectures, tutor students, hold office hours, and frequently teach independent courses in which the professor is seldom seen (Bailey 1982).

This variability in the TAs' role contributes significantly to the difficulties foreign teaching assistants (FTAs) encounter in attempting to be successful teachers in American universities. In the 1970s TA training programs began to emerge. By the end of the decade of the 80s the trend had become something of a movement with a shift in emphasis from TAs in general to FTAs in particular.

Bailey (1983) coined the phrase "foreign TA problem" to encompass the wide-spread dissatisfaction of students, parents, faculty and administrators with the quality of teaching provided to undergraduates by FTAs. The FTA problem is far from being resolved even into the decade of the 90s, although a significant amount of research has provided new insights. Numerous descriptive studies support the idea of training programs for TAs (Barrus, Armstrong, Renfrew, & Garrard, 1974; Buckenmeyer, 1972; Dalgaard, 1982; Davis, 1984; Dege, 1981; Donato, 1983;

Garland, 1983; Goepper & Knorre, 1980; Muhlestein & De Facio, 1974; Royal, 1972; Russo, 1982; Siebring, 1972).

The "FTA problem" is threefold, consisting of linguistic, cross-cultural and pedagogical factors. Investigators reporting difficulties of FTAs in communicative competence state that FTAs lack proficiency in speaking the English language which seriously effects their teaching performance in the classroom environment (Avsar, 1980; Bailey, 1983; Bailey, 1982; Davis, 1984; Gottschalk, 1985; Heller, 1985a; Swanbeck, 1981, and The Chronicle of Higher Education, 1984). These studies urge FTAs to participate in an on-going TA training program.

The linguistic component of the FTA problem is enormously complex. Ard (1987) claims that little attention has been given to uncovering the relationships between second language acquisition theory and how FTAs acquire or fail to acquire language abilities necessary to succeed in teaching undergraduate courses at the university level. Bailey, Pialorski, and Zukowski-Faust (1984) and Rounds (1987) concur with this point that research in second language acquisition is greatly needed to determine: (a) how FTAs master the language skills required in their positions; (b) how colleges can determine which of these foreign students do not have sufficient language abilities; and (c) how colleges can aid those foreign students who are lacking in both basic communications and cross-cultural skills.

Evidence exists that language may not necessarily be the root cause of the FTA problem. Cultural differences are thought to be largely responsible for the classroom difficulties experienced by both U.S.-born students and their foreign teachers. Bernhardt (1987) notes that FTAs and U.S.-born students do not share similar educational experiences, nor are they socialized toward education in a similar manner.

Shirvani (1987) investigated the differences in communications styles of native and non-native teaching assistants which were statistically significant at the .05-level between these groups. Native students perceived non-native teaching assistants as being less relaxed (not calm but nervous under pressure), did not encourage students to do well in their studies, not friendly but contentious, and generally poorer communicators than perceived of native TAs. The investigator attributed the differences student perceptions to non-native TAs.

Impressions left by the TAs both native and foreign are long lasting. Barlow (1985) found in his investigation that students' first impressions of teaching assistants remained the same at the beginning, middle and end of the semester. Providing feedback to TAs during the semester to improve their teaching did not result in more favorable overall student perceptions of effective teaching at the end of the semester. Teaching behavior improved, as perceived by TAs themselves and their raters, but it was not reflected in student evaluations because of the effect of primacy (evaluations made early and late in a course are highly associated) and the strength of the initial perceptions.

Specific instructional dimensions that did not change were communications skills, knowledge of subject matter, clarity of presentation, and warmth and concern-general for students did not change. These characteristic were judged more quickly by students than other instructional dimension where significant changes in student perceptions were noted in this study. Those instructional qualities that did change were TA enthusiasm, warmth, and concern for students.

Research shows that male and female students rated both male and female TAs (native and foreign) the same. In the Barlow (1985) study gender of the TA and gender of the student was not a factor influencing student perceptions of the teaching effective of foreign and native TAs. Although, age was a factor, older TAs received significantly lower ratings than younger TAs; this was especially noted for the two oldest TAs who were 35 years of age. A gender preference for one TA gender over another has been noted for full-time faculty in the present study and in other research investigations. Feldman (1987) argues that the age of most TAs is substantially closer to that of the students than noted for most full-time faculty, hence closer in overall perspectives.

An FTA will enter an American classroom with a set of expectations about teaching and anticipated student behaviors that often differ from those expectations held by their students. The potential for conflict and disruption of the educational process is therefore high.

The former director of Yale's Artificial Intelligence Laboratory (AIL), Roger Schank (1984), believed that empathy is the highest form of understanding, a form which can only develop through shared experiences. Constantinides (1987) speculated that this is perhaps the reason why native, English speaking TAs enter American higher education classrooms with a higher probability of success than FTAs because the native TAs will have a set of expectations that are similar to those of American students, in other words, with a higher level of empathy.

Kleinfeld, McDearmid, Grubis, and Parrett (1984) in a study conducted on the teacher preparation program at the University of Alaska at Fairbanks noted that affective qualities of rapport, concern and empathy were rated substantially higher among students, faculty, and parents than any other "effective teacher" quality. Constantinides (1987) argued that the ability to establish rapport with culturally different students by showing genuine concerns for student learning and intellectual development is the first step toward developing empathy for American institutions and classroom pedagogical dynamics. Her article mentioned three important aspects in the promotion of effective teaching. These include a knowledge of the:

1. general expectations based on the philosophy of education in American culture;
2. specific expectations of the FTA's institution or academic community; and
3. the expectations of the discipline in which the FTA is expected to teach (p. 5).

Understanding the philosophy and expectations of American education is essential to the successful teaching of FTAs. Equally important is knowledge and skill of classroom ecology. The importance of effective classroom management is not a new idea. Kounlin (1970) was one of the first researchers to establish the fact that a well-managed classroom is not a discrepant event. Consistent teacher efforts to create and maintain conditions that enhance student learning is the trade mark of an effective teacher.

The probability that FTAs will enter the teaching arena with this ability is unreasonable to expect. However, it is not unreasonable to expect FTAs to develop these skill early in their teaching careers, if guidance is provided by those who are qualified (teach, research, and publish regularly) to evaluate those involved with instruction.

The National Study of Teaching Assistants (1987) recommends that foreign students spend at least one academic year in the United States before they assume formal teaching responsibilities. One term of casual contact is not sufficient to counterbalance a lifetime of expectations. Constantanides makes the following observation:

> If the FTAs are not made aware of the major differences between the American undergraduates and themselves as undergraduates, then they will have difficulty communicating effectively in a classroom. If they employ their own educational backgrounds to find appropriate teaching behaviors and styles, then the American students will find the methods unfamiliar, uncomfortable, and unhelpful, causing complaints. (p. 5)

Student Assessment of Teaching Instrument

Evaluation of teaching effectiveness is of particular interest to three groups of decision-makers: first, administrators responsible for evaluating faculty with respect to tenure, promotion, and retention; second, faculty who desire feedback on teaching performance to facilitate self-improvement and student learning/achievement; and third, students who seek factual information to aid in the selection process of both instructors and courses (Wotruba & Wright, 1974). There is no paucity of student evaluation instruments. According to Seldin (1984) virtually every institution of higher education across the nation utilizes some form of student evaluation specific to that institution. This fact alone serves to verify the interest which institutions have shown by the inclusion of student input in formative and summative evaluation processes.

Whether the student evaluation instrument employed at Oregon State University is the best available is not an issue. At the present time it is the instrument designed to meet the needs of our academic community, and the instrument adopted by the central administration and faculty to evaluate effective teaching.

This is not to imply that further instrument development is unnecessary. If the chief objective of the OSU's Strategic Plan, "strengthening programs and rewards that promotes and recognizes good teaching," (p. 7) is to come to fruition, then a consorted effort must be made to develop valid and reliable means of conducting faculty evaluations. A factor common to all strategic plans is the continued change that they bring to the university as goals, aims, and objectives are identified and operationalized to achieve new institutional missions.

It is generally understood that standards of effective teaching will vary from one institution to another. Similarly, the criteria on which these standards are based will also display enormous variation and will also tend to invite heated polemics.

The lack of uniformly accepted external criteria of teaching effectiveness is the primary reason why validation studies of a given instrument are seldom conducted or, if conducted, are often not accepted by all units at a university. In spite of the variability of evaluation criteria, the final content of most student evaluation instruments nation-wide is remarkably similar (Aleamoni, 1987).

When systematic evaluation systems are employed, the instrument (content, format, and procedures for administration) should be based on research data from a review of relevant literature. The results of instruments used at other institutions that rely solely on pencil and paper techniques to measure effective teaching, is based on informed opinion and positive attitudes of the members of the drafting team. The specific methodology used in the development of the Student Assessment of Teaching Instrument used at OSU was largely derived from the five-step research methodology of Wotruba and Wright (1974). These steps include the following:

<u>Step 1</u>: Development of an item pool involved a survey of each unit at the university for faculty input on evaluation strategies and the selection of those items most reflective of effective teaching in a particular unit. Part two of the item pool development consisted of a search of the literature base and a survey of a large number of student evaluations of effective teaching forms from other sources (see Appendix B). From all of these sources items that comprise the final form met the following criteria:

a. provide measures of effective teaching as viewed by various committee members involved in the selection process;
b. avoid redundancy and ambiguity in the instrument, those items which measured only collinear dimensions of effective teaching;
c. represent, at least in part, the teaching situations in all colleges and departments campus-wide;
d. facilitate respondent accuracy and speed in the completion of the rating form;
e. provide tangible results that could form the bases of promotion, tenure, and retention decisions; and

f. create a data base that could be used as a starting point for a formative evaluation process.

Step 2: The screening of the item pool. This process was essential as a means of initially reducing the number of candidate items to a manageable level. Items where placed into six categories:

1. attitude toward students;
2. presentation strategies;
3. personal attributes;
4. class atmosphere;
5. course mechanics; and
6. attitude toward the subject.

The categories above were derived from analysis of current investigations published in refereed journals. Items not perceived as fitting one of the six classifications were placed into a miscellaneous or "other" category for later consideration. Following the initial sorting procedure, nine committee members rated each item and category (most, moderate, and least important) based on the degree of acceptance among the committee members. An equal number of items were included in each category.

Step 3: Analyis of categories and composite items. The six categories and composite items listed above were analyzed to determine their ability to discriminate between components of effective teaching and student ability to rate instructors accurately on each component with a five-point scale.

The following criteria were used in the analysis:

1. each category should contain only a few items to avoid boredom and/or a halo effect;
2. only items clearly defined by a category and accepted by the majority of committee members would be included;

3. items unable to be placed in one of the six categories even though regarded by a majority of committee members as "most important" would be retained to determine their relationship to items more clearly categorized; and
4. items would be included that were previously identified by administration and faculty as being the most reflective of effective teaching.

Step 4: Faculty and students on the development committee completed a response analysis. Construction of the instrument began with a rank ordering of each item or factor as to their importance to committee members/faculty and on student's ability to rate instructors accurately with each item. This non-parametric process involved testing the agreement between faculty and student groups on these two factors. The results of the analysis would show how items group into factors, and if these faculty and student factor groupings matched those generated in step two. The factor loadings thus generated could be used in step five.

Step 5: Development of an evaluation instrument. A number of alternative approaches in the construction of an evaluation instrument have been identified in the literature base. Five alternatives are listed below:

1. select items on the importance criterion alone;
2. select items on the ability to rate criterion alone;
3. select items on factor loadings alone;
4. select items based on a combination of the above three criteria; and
5. select items based on two or three of the above criteria, with whatever weighing system is considered appropriate.

Alternative number four is the most comprehensive. The use of the importance criterion alone in this study is based on the rationale that items selected would reflect what committee members believed important in assessing effective teaching.

CHAPTER III

DESIGN AND METHODOLOGY

Introduction

The purpose of this study is to analyze and interpret the results of fall term student ratings of faculty teaching performance (n=40,339) in order to determine whether relationships exist between a range of demographic variables and the ratings faculty receive from students enrolled in their classes. Data analysis was conducted using the Statistical Package for the Social Sciences (SPSSX) on the IBM 4381. The five percent probability level was chosen as the significant level for this study. Specifically, this investigation centered upon the following research questions:

1. Is there a difference in how students evaluate faculty across major and non-major courses?
2. Is there a difference in how students evaluate faculty across required versus elective courses?
3. Is there a difference in how students evaluate faculty based on overall GPA?
4. Is there a difference in how students evaluate faculty based on expected grade?
5. Is there a difference in how students evaluate faculty based on percent of class attendance?
6. Is there a difference in how students evaluate faculty across graduate versus undergraduate courses?
7. Is there a gender bias in the way male and female students evaluate female and male instructors?

8. Is there a difference in how students evaluate faculty based on class size?
9. Is there a difference between course evaluations of tenured and nontenure faculty?
10. Is there a difference between student evaluations of non-international graduate teaching assistants and full-time faculty members?
11. Is there a difference between student evaluations of international graduate teaching assistants and non-international teaching assistants?
12. Is there a relationship between course evaluations and faculty publication rates?

Population

A population of 645 full-time faculty and 273 teaching assistants from 89 departments in 11 academic colleges across Oregon State University were evaluated by the students (n=40,339) in their classes using the Student Assessment of Teaching Instrument fall term 1989. The gender ratio of the population consisted of 55% male and 45% female faculty in the following academic ranks/colleges as shown in Table 1.

The total of 645 faculty in the study represents 98 percent of the faculty contracted at O.S.U. to fulfill teaching responsibilities as a part of their professional duties over the 1989-1990 academic year. The average teaching load at O.S.U. varies from 9-15 hours per year depending on the college analyzed; contrast this figure with 3-5 hours at the University of Oregon, and from 5-9 hours at Portland State University.

The data presented in Table 1 provide the number of faculty and teaching assistants in each academic rank across Oregon State University.

Table 1

Number of Full-Time Faculty and Teaching Assistants by Academic Ranks

Academic Ranks	N	Percent of total
International T.A's	49	5.3
Non-International T.A's	193	21.0
Grad./Research. Assists.	31	3.4
Instructors	128	13.9
Assistant Professors	177	19.3
Associate Professors	163	17.8
Full Professors	177	19.3
TOTAL	918	100.00

Table 2 presents the total number and percent of full-time faculty assigned to each academic unit. A total of 918 full-time faculty (645) and teaching assistants (273) constitute the study population. Teaching assistants were further broken down to include 242 Graduate Teaching Assistants and 31 Graduate Research Assistants (GRA) and Research Assistants Unclassified (RAU). Visiting faculty and guest lectures where not included in the study.

Table 2

Number and Percent of Full-Time Faculty by Colleges

Academic Units	N	Percent of total
Agricultural Sciences	57	9.0
Business	52	8.0
Education	29	4.4
Engineering	81	12.5
Health and Human Performance	40	6.0
Home Economics	30	5.0
Liberal Arts	192	30.0
Oceanography	1	0.1
Pharmacy	8	1.0
ROTC	3	0.5
Science	152	23.5
TOTAL	645	100.00

Table 3 shows one-way ANOVA results comparing overall teaching performance ratings of faculty from eleven colleges. Table 4 shows Newman-Keuls multiple comparison of teaching performance means these same eleven colleges.

Table 3

Student Ratings of Faculty Teaching Performance by Colleges

Source	df	SS	MS	F
Between groups	10	1146.3654	26.9045	187.2318
Within groups	39800	24368.3730	.6123	
Total	39810	25574.7384		

p=.0000

.05 F(10,39811)=1.83

Table 4

Mean Comparisons of Student Ratings of Faculty Teaching Performance by Colleges

College	Means	Student N	Faculty N
ROTC	3.6201 (a) *	102	3
Health & PE	3.4726 (a)	3,609	40
Education	3.4462 (a)	961	29
Agriculture	3.3676 (b)	1,580	57
Home Econ	3.2351 (c)	1,440	30
Liberal Arts	3.1817 (d)	10,594	192
Oceanography	3.1764 (d,e)	70	1
Business	3.0492 (e)	3,482	52
Engineering	3.0250 (e)	3,491	81
Science	2.9835 (f)	13,760	152
Pharmacy	2.7325 (g)	722	8
Overall	3.1229	39,811	645

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Evaluation Instrument

The current Oregon State University Student Assessment of Teaching form was first used fall term 1989. It was a culmination of an enormous amount of time and effort on the part of members of the Committee for the Advancement of Teaching (CFAT). The process began in May 1986 with the Final Report (ad hoc) Committee on Evaluation of Teaching Oregon State University.

This document addressed the objections and criticisms of the academic community regarding the use of the "Orange Card" system of evaluation. A 1986-1987 survey of OSU faculty revealed that 52 percent of the faculty were using some alternative form of student evaluation than the orange punch card system. Objections ranged from having little relevance to a particular discipline to the obvious ambiguities in the way questions were stated.

Others criticized the predominance of educational jargon in the questions constructed. Many questioned whether, for example, uses a wide range of media presentation, had practical significance to any other discipline beyond education. A search was soon initiated for an evaluation system specifically designed for campus-wide application. The committee, chaired by the O.S.U. President, set a goal of providing recommendations that would improve teaching and administrative decision making processes through the use of student evaluations.

The Student Assessment of Teaching Instrument (SATI) is composed of two parts and three sections (see Appendix A). In **Part I** a computerized (pencil in the dot) rating form consists of two sections. **Section 1** asks for information to evaluate and to improve instruction. This section is composed of a 12-item format with a five-point Likert scale ranging from strongly **DISAGREE** (1) to **STRONGLY AGREE** (5), and a separate column designated as **NO BASIS FOR OPINION**.

Section 2 is a seven-item format asking for demographic data about gender of student and instructor, whether the course is required or an elective and if the course is included in a major/non-major area. The section also asks the percent of class attendance of the student, the overall student's GPA, class status, and the grade the student thinks he/she will receive at the end of the course. **Part II** of the instrument consists of an attached sheet asking for students' written comments about the instructor on any of the items of Part I, especially those items that receive a below average score.

In Part II, **Section 3**, of the evaluation form students are made aware of the option to have placed in a particular instructor's file any written comments signed by the student and addressed to the appropriate depart-mental Chair, Head, or Dean. These letters under existing university guidelines constitute "viable" data when an instructor is pending review for the purpose of promotion, tenure, merit pay increases, and the granting of professional privileges.

The guidelines for the distribution and implementation of the Student Assessment of Teaching form are stated in the <u>Rating Instrument Usage Manual</u>. The guidelines recommend the following protocol:

1. students should remain anonymous on the rating instrument;
2. if the instructor distributes the rating instrument, she/he should leave the classroom and either collect them later, or have a student or an assistant gather the ratings;
3. administration should be during the last two weeks of the term and preferably not immediately after the final exam;
4. students should be given ample time to complete both sections of the evaluation form; and
5. students should be encouraged to take the time to fill out the evaluation form with care and accuracy.

Data Gathering with SATI

This instrument consists of 12 items grouped into four scales which the respondent answers by using a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree see Appendix A). The 12 items on SATI are listed below as they appear on the instrument:

1. Course objectives and requirements were clearly presented to me.
2. The instructor was well prepared and organized.
3. The instructor explained the material clearly.
4. The instructor was sensitive to my/the class ability to understand the material.
5. The instructor stimulated enthusiasm for the subject matter of the course.
6. The instructor provided scheduled office hours or was readily available for consultation with me.
7. The instructor was fair and impartial in dealing with me.
8. The instructor encouraged me to think for myself.
9. The examinations were relevant to the reading assignments and to the material presented in class.
10. The instructor used good communication skills.
11. As a result of having this instructor, I have learned a significant number of new ideas and/or skills.
12. All things considered, I was favorably impressed by this instructor.

Table 5 illustrates the internal consistency of SATI to accurately measure student perceptions to effective teaching. The results of four consecutive terms on student ratings of faculty teaching performance with SATI have been recorded. The range in the overall means for fall, winter, and spring terms was less than 0.05 on a five point scale. Summer term overall mean score was higher, but only less than 0.3 difference from the lowest mean of the three academic terms.

The range of mean scores for the demographic variables and levels of the variables constituting the overall mean was less than 0.2 on a five-point scale for the three academic terms. The range for summer term was greater on the demographic variables and levels of the variables. The net result for summer term were higher ratings on all demographic variables. Why summer terms rating are higher than other terms is unknown. Future data collection with SATI may provide insights into this phenomena.

Data for hypothesis 13 involved tabulation of the number of published articles appearing in refereed journals over a five-year period beginning with the 1985-1986 academic school year and ending with the 1989-1990 academic year. The five year publication period was based on the knowledge that articles sent in for review by professional journals take from one to five years to appear in print. One year was thought to be inadequate to determine publication rates. A span of five years was considered more than adequate to cover the 1989-1990 academic year. The publication rate data for the College of Education alone was compared with five other universities randomly selected from the Pacific Athletic Conference.

From these tabulations publication (research productivity) rates were computed for all instructors who had formal classes as recorded in the OSU schedule of classes for fall term, who were teaching a minimum of six hours per academic year, and who had full-time faculty teaching status. A comparison was then made between the results on the over-all subscale of teaching ratings on the Student Assessment of Teaching Instrument and the publication rate for each instructor meeting the above three criteria.

Table 5

Faculty Ratings Over Four Consecutive Academic Terms

Variable	Fall 89	Winter 90	Spring 90	Summer 90
Class Status				
Major	3.1002	3.1484	3.1329	3.3836
Non-major	3.1531	3.1541	3.2060	3.4478
	0.0529	0.0057	0.0731	0.0642
Course Selection				
Required	3.0712	3.0998	3.1034	3.3366
Elective	3.2401	3.2457	3.2782	3.5172
	0.1689	0.1459	0.1748	0.1806
GPA				
3.50-4.00	3.1660	3.1924	3.2092	3.4411
3.00-3.49	3.1282	3.1393	3.1541	3.3669
2.00-2.99	3.0976	3.1295	3.1561	3.3362
0.00-1.99	3.0465	3.1071	3.0844	3.1633
Expected Grade				
A	3.3088	3.3163	3.3364	3.4963
B	3.0607	3.0829	3.0776	3.2143
C	2.7925	2.8070	2.8280	2.9883
D	2.5256	2.5437	2.6844	1.4500
F	0.0000	0.0000	0.0000	0.0000
Classroom Attendance				
81-100 %	3.1251	3.1698	3.1927	3.4097
61-80 %	2.8934	2.9553	2.9568	3.2108
41-60 %	2.7706	2.8528	2.7793	3.0109
21-40 %	2.6443	2.8134	2.9451	3.1667
0-20 %	2.7682	2.9198	2.8829	3.8182

(Table 5 continued)

Variable	Fall 89	Winter 90	Spring 90	Summer 90
Class Status				
Freshman	3.1167	3.1509	3.1801	3.5979
Sophomore	3.1268	3.1515	3.1560	3.3636
Junior	3.1349	3.1308	3.1506	3.3167
Senior	3.0792	3.1247	3.1559	3.3490
Graduate	3.2490	3.2516	3.2558	3.4375
Gender (Student)				
Male	3.1028	3.1169	3.1314	3.3414
Female	3.1532	3.1918	3.2147	3.4585
	0.0504	0.0749	0.0833	0.1171
Gender (Instructor)				
Male	3.1547	3.1489	3.1502	3.3856
Female	3.2111	3.2570	3.2834	3.4760
	0.0564	0.1081	0.1332	0.0904
Class Size				
1-28	3.3162	3.2194	3.2673	3.4336
29-50	2.9877	3.1346	3.1009	3.3495
51-100	3.0534	3.0263	3.0463	3.4398
100+	3.1331	3.0734	3.0327	3.1223
Overall Means	3.1229	3.1497	3.1667	3.3907

Procedures and Methods of Data Analysis

The study used analysis of variance (ANOVA), multiple analysis of variance (MANOVA), and a Pearson product moment correlation to analyze the data. The goal of the study sought to identify potential relationships between select demographic variables and student ratings of faculty teaching performance on 12 dependent variables constituting the Student Assessment of Teaching Instrument (SATI).

The specific research questions of this study centered on determining if relationships existed between select demographic variables and those ratings instructors received from students in their classes. In order to determine the extent to which select instructor demographic variables were related to student ratings, 13 null hypotheses were posited. The alpha level of significance for each statistical test was established at .05. The null hypotheses follow:

HO:1 There is no significant difference in course evaluations between faculty across major and non-major courses.

HO:2 There is no significant difference in course evaluations between faculty across required versus elective courses.

HO:3 There is no significant relationship between mean faculty ratings and student GPA.

HO:4 There is no significant relationship between mean faculty ratings and student expected grade.

HO:5 There is no significant relationship between mean faculty ratings and student percent of classroom attendance.

HO:6 There is no significant difference in course evaluations between faculty across graduate and undergraduate courses.

HO:7 There is no significant difference in course evaluations between male and female students.

HO:8 There is no significant difference in course evaluations between male and female faculty.

HO:9 There is no significant difference in student evaluations of faculty based on class size.

HO:10 There is no significant difference in course evaluations between tenured and nontenure faculty.

HO:11 There is no significant difference in the course evaluations of non-international graduate teaching assistants and full-time faculty members.

HO:12 There is no significant difference in course evaluations between non-international and international graduated teaching assistants.

HO:13 There is no significant relationship between course evaluations and publication rates among full-time faculty.

Only those fifteen items on the Student Assessment of Teaching Instrument that directly pertained to the stated research questions were evaluated. Analysis of variance (ANOVA) procedure was used to address each of the previously stated research questions. In particular, a one- and two-way ANOVA were used to evaluate full-time faculty and graduate teaching assistants included in hypotheses 1-12, and a Pearson product moment correlation was used for hypotheses 13. In hypothesis 13, faculty publication rates and student evaluations of teaching effectiveness were correlated for the 1989-1990 academic year.

Hypothesis 9 involved determination of a class size designation to investigate the effect of the class size variable on how students evaluate the teaching performance of faculty in the classes in which students are enrolled. This designation was based on the Minimum Class Size Policy set by the Registrars Office at Oregon State University.

The minimum number of students allowed per undergraduate and graduate classes according to the policy are as follows: lower division = 16, upper division = 12, and graduate = 8. Student numbers below these benchmarks represent small classes and above these benchmarks represent large classes. The total number of classes in the study was 786 for the small class size and 947 for the large class size with a total of 1733 classes in the study. For hypotheses 11 student ratings of native teaching assistants (NITA) on the over-all teaching effectiveness subscale were compared with student ratings of full-time faculty.

Similarly, hypothesis 12 compares native and non-native teaching assistants on the over-all teaching effectiveness subscale of the Assessment of Teaching Instrument. The data required to answer research questions one through nine were ascertained by computer analysis of fall term's student rating forms sent by each college to the computer center for analysis.

Hypothesis 13 involved computation of the publication rates for all faculty who met the criteria of having full-time teaching status, who were responsible for a minimum of six hours of instruction per academic year, and who taught fall term of the 1989-1990 academic year. The unit of analysis in this segment of the study was the academic unit (college).

The students, instructors, and classes not participating in the current study were considered as missing data. Collection of data followed guidelines in accordance with Federal and University regulations with respect to the use of human subjects in a study:

1. only faculty with instructional responsibility are eligible to participate in the study; and
2. the names of faculty will be kept in confidence and would not be associated with any data collected or in any publication.

An extensive on line LIRS (Library Information Retrieval System) search was conducted on 645 faculty members. The BRS/After Dark rates were significantly lowered from the original projected cost of $2,800 for the eight to five prime time hours to conduct a full search of faculty publications.

Not all academic disciplines were available in BRS. For example, in the areas of mathematics and the biological sciences, Mathematics Index and Biological Abstracts were necessary to access at a cost exceeding three times the cost of BRS rates. The BRS offers a convenient and cost-effective access to journal information in 150 data bases in the following fields: Medicine, Pharmacology, Psychology and Psychiatry, Arts and Humanities, Education, and Business.

The reams of data acquired for faculty members in this study were meticulously tabulated according to publications in refereed journals as defined by The American Library Association Glossary of Library and Information Science, or TALAGLIS (see definition section p. 9). Journals were identified and recorded as refereed or not. In a number of cases it was necessary to make contact with the editors of a particular journal when refereed status was unknown.

Variation among refereed journals was discovered to be enormous. In phone conversations with the editors of several of the more "prestigious" journals, it was noted that it may take upwards of from three to five years before an article submitted for publication would appear in print. For most refereed journals identified in this study required from one to two years for an article to appear in print.

Articles meeting and/or exceeding the definition of refereed journals provided by TALAGLIS were included in the study. Those publications that did not meet the TALAGLIS criteria of a refereed journal were not counted in the final faculty publication rate. Publications rejected from faculty publication counts included bulletins, technical reports, books, chapters in books, book reviews, refereed abstracts of papers submitted to meetings, government reports, commentaries, ERIC (Education Research Information) documents, etc. The final list of refereed journals meeting TALAGLIS criteria exceeded 2,500.

Publication data acquired for each faculty were carefully compared with college faculty publication compilations. Each college at OSU, with the exception of the College of Education, compiles faculty publication data at the deans office. For example, the College of Science publishes Science Record, the College of Liberal Arts prints the CLA; Colleges of Home Economics, Engineering, Business, and Pharmacy tabulate publication data for use by faculty and administration. Currently, there is no annual compendium or centralize data base containing faculty publications and other essential faculty data across all departments and colleges campus wide. This research endeavor did not have the luxury of a centralized data base to draw from, and for this reason, it was necessary to "merge" three existing data bases by hand for later analysis, at a total cost exceeding $7,300.

The data on publication rates obtained from the LIRS search and college faculty publication compendiums were compared to data acquired from yet a third source, e.g. faculty members themselves. In the original research design of the proposal, faculty publication data would come exclusively from faculty willing to participate in the study. It was estimated that publication data on 250 faculty would be adequate to compare faculty publication rates with their ratings on the Student Assessment of Teaching Instrument using the stratified random sample procedure outlined below:

1. Separation of OSU faculty into academic units (strata) based on the following criteria:
    a. full-time faculty member.
    b. teach a minimum six hours per academic year.
    c. teach fall term of the 1989-1990.

2. Select a sample from each of 13 academic units
   a. based on what percent (Pi, i= 1 to 13) faculty in each academic unit represented in the total faculty meeting the above three criteria.

$$Pi = \frac{\text{number of faculty in unit \#i}}{\text{total faculty number}}$$

   b. compute the sample size (Ni) for each academic unit.

$$Ni = Pi * 250 \ (i=1 \text{ to } 13)$$

   c. apply Simple Random Sampling Method (SRS) to select the samples.
   d. generate random numbers from computer.
   e. repeat the step 2. for each academic unit.

3. Conduct a LIRS search on faculty randomly selected from each academic unit:
   a. contact each instructor to verify the accuracy of the publication rate data.
   b. have each instructor provide a signature verifying the authentic and accuracy of the computer generated list of publications in refereed journals.

Only 60 faculty were contacted before the study was temporarily halted by a concerned dean who wanted more assurance that the confidentiality factor was sufficiently rigorous to protect individuals who volunteered to participate in the study. An investigation by the Vice President of Academic Affairs and members of the Committee for the Advancement of Teaching revealed that sufficient precautionary measures for this type of research was in fact in place. However, due to the volatile nature of this kind of research, and the suspicion and negative attitudes associated with it, contact with faculty was terminated.

The publication data derived from the 60 willing participants were triangulated with publication data from the LIRS search and college publication data. The results indicated a high degree of corroboration between these two data sources. A 98.9% degree of confidence was computed that articles submitted for publication did in fact appear in the computer data bases researched; and that articles appeared in refereed journals meeting TALAGLIS criteria.

Publication rate data from the College of Education at Oregon State University were compared to five other universities in the Pacific Athletic Conference (Pac 10). Publication means for all academic ranks in the College of Education at each of the five randomly selected institution were computed and presented in Table 24. Some university Offices of the Registrar, Admissions, and Academic Affairs provided faculty names, rank, and FTE only, while others provided lists of publications, making it necessary to acquire schedules of classes and to make phone calls to confirm faculty teaching responsibilities of more than six hours.

The Pac 10 Conference consists of the following universities: Oregon State University, University of Oregon, Washington State University, University of Washington, California State University at Berkeley, University of Southern California, Stanford University, University of California at Los Angeles, Arizona State University, and the University of Arizona. Universities randomly selected for this research included the following: 1. Washington State University, 2. Arizona State University, 3. Stanford University, 4. University of Southern California, 5. University of California at Los Angeles.

CHAPTER IV

ANALYSIS OF DATA

Introduction

The study was an attempt to analyze correlation results from fall term, 40,339 student ratings of faculty teaching performance and a range of demographic variables. A series of ANOVA's were used to determine by college statistical differences among course evaluations for the total, undergraduate, and graduate data sets. Comparisons among course evaluations were performed with respect to the following independent variables:

1. Course in Major/Not in Major.
2. Reason for Taking Course (i.e., required/elective).
3. Course Level and Faculty Evaluations
4. Gender of Student and Faculty.
5. Class Size and Faculty Evaluations.
6. Student Percent of Classroom Attendance.
7. Student Grade Point Average (GPA).
8. Student Expected Grade
9. Undergraduate/Graduate Students and Faculty Ratings
10. Native/International Graduate Teaching Assistants.
11. Faculty Tenure Status (i.e., tenured/nontenure).
12. Faculty Publication Rate and Evaluation Scores.

A Pearson product moment correlation statistic was used to determine statistical differences between course evaluations and publication rates among full-time faculty. Both general and specific research findings are presented below.

General Demographic Variable Results

Statistically significant relationships were found between course evaluations and all demographic variables with exception of faculty tenure status, non-international teaching assistants/full-time faculty teaching performance, and class size or the number of students enrolled in each course taught. Table 3 provides an overall course mean evaluation results of full-time faculty members and graduate teaching assistants on select demographic variables chosen for this study. Findings, from the analysis of publication rate data and overall scores on faculty course evaluations, indicated that a low negative correlation coefficient was obtained when faculty publication rate data and faculty evaluation mean scores were compared.

Specific Findings From Demographic Variables

The results of the data analyses are presented with respect to their appropriate hypotheses. In hypothesis #11 other than the group designated International Teaching Assistants (ITA), all academic rank mean score differences were less than 0.06 on a five point scale. Only those mean differences greater than or equal to 0.3 were considered to be of practical significance. In Hypothesis #12, Non-International (NITA) and International (ITA) Teaching Assistants were compared on the bases of differences in teaching performance on 12 subscales of the Student Assess-ment of Teaching Inventory (SATI).

On all subscales the NITA group received the lowest ratings on subscale #9 (examinations were relevant to assignments and lecture). The highest ratings for this group was on subscale #10 (The instructor used good communications skills) with one exception; the freshman students expressed that course objectives and requirements were clearly presented (subscale #1) as NITA's strongest point.

Freshman and sophomores felt the same way about the ITA group on subscale #1. However, junior level students rated the ITA group highest on their having learned a significant number of new ideas (subscale #11), while the seniors felt the ITA group's strongest point was good communica-tions skills (subscale #10). Overall differences in mean scores between ITA and NITA groups was 0.3341 on a five point scale.

At the college level unit of analysis, the College of Home Economics showed a moderately positive correlation coefficient of .4488, and the College of Pharmacy with a moderately negative correlation of -0.3143 between faculty publication rates and course evaluation mean results. The College of Liberal Arts showed a high negative correlation coefficient of -.9635 between the above variables.

In hypotheses #7 and #8 gender of the student and instructor interacted with nine of the twelve subscales on SATI. At the college level of analysis, Engineering and Home Economics were compared, and the Colleges of Agriculture and Education on the gender variable. There was a statistical difference in the way male and female students rate male and female instructors in the aforementioned colleges.

Male and female students rated male instructors higher than female instructors in the College of Home Economics, but the reverse was true in the College of Engineering. Female instructors had a slight advantage over their male instructor counterparts in overall ratings for all colleges combined, a difference that is significant at the .05 level of confidence.

A three-way ANOVA was conducted on GPA (hypothesis #3), expected grade (hypothesis #4), and classroom attendance (#5). A near linear relationship existed between the levels of each of these variables and the way students rate instructor teaching performance. Students who attended class the highest percent of the time throughout their courses also expected the highest grades, reported the highest GPA, and gave among the highest instructor ratings.

The highest faculty ratings were given by students auditing (H) a course and the A grades, the means of both were statistically indistinguishable. The third highest mean rating scores were given by the F grade designate, reasons for these data anomaly are unknown.

A significant difference was found for student ratings of faculty teaching performance in major/non-major (hypothesis #1) and required/elective (hypothesis #2). Both the non-major and elective courses received significantly higher ratings on SATI.

For hypothesis #6 significant differences were found between undergraduate and graduate student ratings of faculty teaching performance. The ANOVA results at the course level of analysis showed that the freshman level in the undergraduate student grouping and graduate students gave the highest overall mean ratings of faculty teaching performance. However, by the time freshman become seniors the difference in rating means between ITA and NITA groups became highly noticeable, ITAs were rated significantly lower than their NITA counterparts.

Class size, hypothesis #9, had no effect on overall faculty rating results. However, at the instructional dimension level (subscales) of SATI students in large classes rated faculty as high, or higher than small class. Faculty in large classes were expected to be better prepared/organized, be clearer in presenting course expectations, and test over relevant material. Students may have perceived that they had less opportunity for interaction in large class sizes, therefore structure was considered important. In small classes, students rated faculty significantly higher on faculty sensitivity to students needs to understand course material, and for student/faculty contact. For hypothesis #10 no significant difference were found between faculty ratings and tenure status; however, significant differences were found between tenure status and publication rates (hypothesis #13).

Data Findings Related to the Null Hypotheses

Hypothesis 1

Hypothesis 1 predicted that there would be no significant difference in how students evaluate professors across major and non-major courses as measured by the Student Assessment of Teaching Instrument. As shown in Table 6, a statistically significant difference $F(1,39237) = 42.2278$, $p=.0000$) was found on overall ratings between major and non-major groups of students; therefore the null hypothesis was rejected.

Table 6

Major and Non-Major Student Ratings of Faculty Teaching Performance

Source	df	SS	MS	F
Between groups	1	26.9045	26.9045	42.2278
Within groups	39235	25115.6094	.6371	
Total	39236	2142.5139		

$p=.0000$

$.05\ F(1,39237)=3.84$

A total of 39,237 out of 40,339 student rating responses were included in the major/non-major (class) variable analysis, with 1,102 (2.8%) missing cases or responses failing to indicate class standing on SATI. Majors constituted 21,801 (56%) of the responses, and 17,436 (44%) for the non-major group. The overall mean evaluation by students in their major was 3.1002 and 3.1531 for non-majors with a difference between groups of 0.0529 a point on a five-point scale.

Table 7 provides a Newman-Keuls multiple comparison of student evaluations of faculty teaching performance means across colleges for the major and non-major categories on the SATI. The ROTC (a) received the highest mean ratings of all colleges, 3.5833 for major and 3.6228 for non-major variables. The next highest mean ratings were those in the Colleges of Education, Agriculture, and Health & P.E. (b), grouped together as being statistically the same.

Table 7

Major and Non-Major Student Ratings of Faculty Teaching Performance by Colleges

	Major		Non-major	
Colleges	Means	N	Means	N
ROTC	3.5833 (a) *	7	3.6228	95
Education	3.4276 (b)	667	3.4732	276
Agriculture	3.3453 (b)	1083	3.4187	468
Health & PE	3.3198 (b)	817	3.5174	2751
Liberal Arts	3.2679 (c)	3766	3.1352	6703
Home Econ	3.2608 (c)	956	3.1920	468
Business	3.0626 (d)	2856	2.9822	585
Engineering	3.0315 (d)	2981	2.9954	450
Science	2.9894 (d)	7986	2.9751	5553
Pharmacy	2.7260 (e)	676	3.0161	23
Oceanography	2.7197 (e)	6	3.2192	64
Overall	3.1002	21801	3.1531	17436

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

The means of Liberal Arts and Home Economics (c) were also grouped together, as were the means of Business, Engineering, and Science (d). The lowest mean ratings were received by the Colleges of Science, Pharmacy and Oceanography on the major class variable with a mean of 2.8117; and the Colleges of Business, Engineering, and Science on non-major class variable with a mean of 2.9842. Both means were below the mean ratings for all colleges combined.

The differences in mean ratings between major and non-major variables within colleges was less than 0.1 for seven of the eleven colleges represented in this study, e.g. ROTC, Education, Agriculture, Home Economics, Business, Engineering, and Science. The Colleges of Health/P.E. and Liberal Arts revealed mean ratings of less than 0.2 on a five-point scale. The lowest overall mean differences for major verses overall mean for non-major variables were in the Colleges of Pharmacy, a mean of 0.2901 and Oceanography 0.4995. Two-way ANOVA was conducted comparing the Colleges of Education and Pharmacy on 12 subscales of SATI.

Between colleges in the major course variable the overall rating scores ranged from a high in ROTC of 3.5833 to a low in Oceanography of 2.7197 with a mean difference of 0.8636. In the non-major course variable the overall rating scores ranged from a high in ROTC of 3.6228 to a low in Science of 2.9751 with a mean difference of 0.6477 on a five-point scale.

Two-way ANOVA was conducted on 27,400 student responses comparing faculty ratings based on student gender and major/non-major courses (Table 8). Results revealed no interaction on 11 of the 12 subscales of SATI. A main effects was discovered on the gender of the student variable for Subscale #7 (The instructor was fair and impartial in dealing with me) which was significant at the 0.05 level of confidence. Female students gave female faculty a mean rating of 3.7 and for non-major a rating of 3.2 points on a five-point scale. A mean difference of 0.5 was noted on this subscale rated by female students.

Table 8

Mean Ratings on 12 Subscales of SATI as a Function of Student Gender and Major

	Male student		Female student	
Variable	Major	Non-major	Major	Non-major
Subscale 1	3.2	3.2	3.1	3.2
Subscale 2	3.3	3.3	3.2	3.3
Subscale 3	3.0	3.0	3.0	3.1
Subscale 4	3.0	3.0	3.1	3.2
Subscale 5	3.2	3.0	3.1	3.1
Subscale 6	3.2	3.3	3.1	3.4
Subscale 7	3.4	3.3	**3.7***	3.2
Subscale 8	3.3	3.2	3.2	3.2
Subscale 9	3.2	3.2	3.1	3.2
Subscale 10	3.2	3.1	3.1	3.2
Subscale 11	3.2	3.0	3.1	3.0
Subscale 12	3.2	3.1	3.1	3.2
Overall				
Major	3.2	3.1	3.2	3.1
Non-major	3.2	3.2	3.1	3.1

* Statistically significant at the $p < .01$. level of confidence.

Hypothesis 2

Hypothesis 2 predicted that there would be no significant difference in how students evaluate faculty across required versus elective courses as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results (Table 9) indicate a statistical difference (.05 F1,39296= 3.84 < 378.5043, p=.0000) was found between the required and elective courses, therefore the null hypothesis was rejected.

Table 9

Faculty Teaching Performance in Required and Elective Courses

Source	df	SS	MS	F
Between groups	1	239.0506	239.0506	378.5043
Within groups	39294	24932.3523		
Total	39295	25171.4029		

p=.0000

.05 F(1,39296)=3.84

A total of 39,296 out of 40,339 student rating responses were included in the required/elective (class) variable analysis, with 1,043 (2.7%) missing cases or responses failing to indicate course selection on SATI. Required classes constituted 27,047 (69%) of the responses, and 12,249 (31%) for the elective group (Table 10).

The overall mean evaluation from students taking required courses was 3.0712, and 3.2401 for those taking elective courses, with a difference between groups of 0.1689 on a five-point scale. Between colleges in the required course variable the overall rating scores ranged from a high in ROTC of 3.5329 to a low in Pharmacy of 2.6475 with a mean difference of 0.8854. In the elective course variable the overall rating scores ranged from a high in ROTC of 3.6890 to a low in Science of 3.0459 with a mean difference of 0.6431 on a five-point scale.

Table 10

Mean Comparisons of Faculty Teaching Performance in Required and Elective Courses by Colleges

	Required		Elective	
Colleges	Means	N	Means	N
ROTC	3.5329 (a)	47	3.6890	54
Education	3.4262 (b)	666	3.4777	282
Health & PE	3.3541 (b)	1,201	3.5323	2355
Agriculture	3.3434 (b)	1,104	3.4266	446
Home Econ	3.1925 (c)	1,025	3.3549	400
Liberal Arts	3.1901 (c)	5,345	3.1750	5,136
Oceanography	3.1430 (c)	13	3.2051	56
Business	3.0365 (d)	3,076	3.1521	372
Engineering	3.0085 (d)	2,846	3.1227	594
Science	2.9707 (d)	11,114	3.0459	2469
Pharmacy	2.6475 (e)	610	3.3566	85
Overall	3.0712	27047	3.2401	12249

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Newman-Keuls multiple comparison of student evaluations of faculty teaching performance means across colleges is provided for the required and elective categories on SATI. The ROTC (a) received the highest mean ratings (3.5329) of all colleges, 3.5329 for required and 3.6890 for elective variables. The next highest mean ratings were given by students in the Colleges of Education, Health & P.E., and Agriculture (b), which are grouped together as being statistically the same.

The means for Liberal Arts, Oceanography, and Home Economics (c) were also grouped together, as were the means of Business, Engineering, and Science (d). The lowest mean ratings were received by the Colleges of Science (2.9707) and Pharmacy (2.6475) on the required class variable. Ratings from both colleges including Business and Engineering were below overall mean for all colleges combined.

The differences in mean ratings for required and elective variables within colleges was less than 0.1 for five of the eleven colleges represented in this study, e.g. Education, Agriculture, Liberal Arts, Oceanography, and Science. The Colleges of ROTC, Health & P.E., Home Economics, Business, and Engineering showed mean ratings of less than 0.2 on a five-point scale. The mean difference between required and elective variables was 0.7091 in the College of Pharmacy.

Two-way ANOVA results comparing major/non-major variables with required/elective variables showed no interaction on any of the 12 subscales of SATI. However, main effects were revealed on seven subscales due to the required/elective course variable with mean ratings significantly higher for faculty teaching the elective courses.

The subscales with main effects were: #2 (The instructor was well prepared and organized), #3 (The instructor explained the material clearly), #4 (The instructor was sensitive to my/the class' ability to understand the material), #5 (The instructor was stimulated enthusiasm for the subject matter of the course), #7 (The instructor was fair and impartial in dealing with me), #9 (The examinations were relevant to the reading assignments and to the material presented in class), and #10 (The instructor used good communication skills).

Main effects due to some other variable other than the required/elective and major/non-major were found on subscales: #1 (Course objectives and requirements were clearly presented to me), #6 (The instructor provided scheduled office hours or was readily available for consultation with me), #8 (The instructor encouraged me to think for myself), #11 (As a result of having this instructor, I have learned a significant number of new ideas and/or skills), and #12 (All things considered, I was favorably impressed by this instructor). Faculty teaching non-major/elective courses received mean ratings significantly higher than faculty teaching major/required courses on all five of these subscales.

Hypothesis 3

Hypothesis 3 predicted that there would be no significant difference between mean faculty ratings by student grade-point average (GPA) as measured by the Student Assessment of Teaching Instrument. The ANOVA results (Table 11) show that a statistical difference (.05 F1,29525=2.60 < 6.9568, p=.0000) was found between 4-levels of GPA, therefore the null hypothesis was rejected.

Table 11

Student Mean Rating Scores of Faculty Teaching Performance by Expressed GPA

Source	df	SS	MS	F
Between groups	3	13.83082	4.6101	6.9568
Within groups	39521	20573.3346	.6627	
Total	29524	20587.1647		

p=.0000

.05 F(3,29525)=2.60

A total of 29,525 out of 40,339 student rating responses were included in the GPA (student) variable analysis, with 10,814 (26.8%) missing cases or responses failing to indicate GPA on the SATI. The GPA variable consisted of four levels groups: Group #1, the 3.50-4.00 range with 136 responses (.5%) of the total; Group #2, 3.00-3.49 with 12,550 (42.5%); Group #3, 2.00-2.99 with 10,537 (36%); and the 0.00-1.99 range or Group #4 with 6302 (21%) of the student responses.

The overall mean evaluation for all groups combined was 3.1245, and separately: Group 1 = 3.1660, Group 2 = 3.1282, Group 3 = 3.0976, and Group 4 =3.0469. A mean difference of 0.1195 on a five-point scale was noted between the lowest mean, group #4 of 3.0465 and the highest mean of 3.1660 for Group #1.

Table 12. lists the Newman-Keuls multiple comparison of means for each level of GPA on SATI. The range in the means of Groups 1-4 was less than .2 on a five-point scale. The 3.50-4.00 GPA (Group 1) gave the highest faculty teaching performance mean rating of 3.1660. Followed by a mean of 3.1282 for the 3.00-3.49 (Group 2) GPA level, which was also statistically indistinguishable from Group 1. The next highest mean rating value was given by GPA level 2.00-2.99 (Group 3) which was not statistically different from the 0.00-1.99 GPA (Group 4).

The relationship between GPA level 0.00-4.00 and the corresponding mean ratings teaching performance was positive. The overall mean rating value for all GPA level was 3.1245. Only the 3.50-4.00 and the 3.00-3.49 GPA levels were above this mean and then only slightly (0.0415).

Table 12

Mean Comparisons of Faculty Teaching Performance by Student Grade Point Average (GPA)

Groups	GPA	Means	N
1	3.50-4.00	3.1660 (a)	136
2	3.00-3.49	3.1282 (a)	12550
3	2.00-2.99	3.0976 (b)	10537
4	0.00-1.99	3.0465 (b)	6302
Overall		3.1245	29,525

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Table 13 shows the results of two-way ANOVA comparing GPA and expected grade of 29,525 students responses that marked both student variables on SATI. No interaction was discovered in the ANOVA. A positive relationship was found in the comparison of these two student variables at all levels. However, nominal differences did occur at GPA level 3.00-3.49 where the F expected grade mean ratings was .07 of a point higher than the D grade level of the expected grade. Similarly, in the 3.5-4.00 GPA level, the D grade level mean was higher than the C grade mean by .18 of a point on a five point scale. These difference were not statistically significant at $p < .05$ level of confidence.

Table 13

Student Mean Rating Scores of Faculty Teaching Performance by Expressed GPA and Expected Grade

Variable	Levels	Means	N
GPA	0.00-1.99	3.0465	136
EXP.GRADE	A	3.4960	33
EXP.GRADE	B	3.0967	39
EXP.GRADE	C	2.8781	43
EXP.GRADE	D	2.6275	12
EXP.GRADE	F	2.5438	9
GPA	2.00-2.99	3.0976	12550
EXP.GRADE	A	3.3723	3728
EXP.GRADE	B	3.0941	5783
EXP.GRADE	C	2.7938	2783
EXP.GRADE	D	2.5092	237
EXP.GRADE	F	2.1005	19

(Table 13 continued)

Variable	Levels	Means	N
GPA	3.00-3.49	3.1282	10537
EXP.GRADE	A	3.2934	4810
EXP.GRADE	B	3.0493	4583
EXP.GRADE	C	2.7617	1085
EXP.GRADE	D	2.5271	50
EXP.GRADE	F	2.5965	9
GPA	3.50-4.00	3.1660	6302
EXP.GRADE	A	3.2626	4385
EXP.GRADE	B	2.9795	1686
EXP.GRADE	C	2.6843	207
EXP.GRADE	D	2.8632	17
EXP.GRADE	F	2.5952	7
Overall	All	3.1229	29,525

Hypothesis 4

Hypothesis 4 predicted that there would be no significant difference between mean faculty ratings by student expected grade as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results (Table 14) indicate a statistical difference (.05 $F1,37509=2.37 < 196.6342$, $p=.0000$) was found between 7-levels of grades, therefore the null hypothesis was rejected.

Table 14

Mean Rating Scores by Student Expected Grade

Source	df	SS	MS	F
Between groups	7	989.8826	123.7353	196.6342
Within groups	37502	19782.8851	.6293	
Total	37508	20843.5481		

$p=.0000$

.05 $F(4,37509)=2.37$

A total of 37,509 out of 40,339 student rating responses were included in the expected grade (student) variable analysis, with 2,830 (7%) missing cases or responses failing to indicate their expected grades on the SATI. The expected grade variable consisted of nine levels, of these only seven were included in the analysis. Each is listed with the corresponding number of student responses and the percentage that number represents for all levels combined. The five letter grades with corresponding numerical values associated with the grade are: A, with 16,549 (44.1%) of the student responses; B, with 15,080 (40.2%); C, with 5,302 (13.2%); D, with 474 (1.3%), and the F with 62 (.2%) of the student responses.

The letter designates without numerical values associated with them are the H or audit with 42 (1%), I or the incomplete grade designate, the G or unsatisfactory/no pass with 44 (0.1), and the E grade designate with 43 (0.1%) of the total student responses. For the purpose of this research only the means of the H, or audit grade of the non-numerical grades was considered in relationship to the letter grades because it had received the highest overall mean rating among the grade designates.

Table 15

Mean Comparisons of Faculty Teaching Performance by Expected Grade

Expected Grade	Overall Means	N
H AUDIT	3.5198 (a) *	42
A	3.3088 (b)	16549
B	3.0607 (c)	15080
C	2.7925 (d)	5302
D	2.5256 (d)	474
F	2.4396 (e)	62
Overall	3.1245	37,509

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Table 15 shows the Newman-Keuls multiple comparison of means for each category of expected grade on the SATI. The H or Audit and A grades (b) gave the highest mean ratings (3.5198 and 3.3088, respectively) for faculty teaching performance, followed by the B grade with a mean rating value of 3.0607. Next highest was the C grade with a mean value of 2.7925. The C and D grades were statistically indistinguishable from each other, giving a mean rating value of 2.6590. The F grade was statistically distinguishable from all other grade levels.

The relationship between letter grades (A-F) and faculty mean rating scores for effective teaching was positive. The overall mean for all grade categories combined was 3.1245. Only the A and H (Audit) grades were above this mean.

Hypothesis 5

Hypothesis 5 predicted that there would be no significant difference between mean faculty ratings and student percent of classroom attendance as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results (Table 16) indicate that a statistical difference (.05 $F4,31462=2.37 < 77.5090$, $p=.0000$) existed between 5-levels of attendance, therefore the null hypothesis was rejected.

Table 16

Student Mean Rating Scores by Percent of Class Attendance

Source	df	SS	MS	F
Between groups	4	203.8826	50.7353	77.5090
Within groups	31457	20640.8851	.6561	
Total	31461	20843.5481		

$p=.0000$

.05 $F(4,31462)=2.37$

A total of 31,462 out of 40,339 student rating responses were included in the classroom attendance (student) variable analysis, with 8,877 (22%) missing cases or responses failing to indicate percent of classroom attendance on SATI. The classroom attendance variable consisted of five levels or ranges in student percent of classroom attendance during the quarter. Each level is listed with the corresponding number of student responses and the percentage that number represents for all levels combined.

The range in Group 1 with 151 (0.5%) students reported to have attended class between 0-20% of the time for fall quarter 1989. Group 2 ranged from 21-40% with 120 (.4%) of the total number of student responses. The range for Group 3 was 41-60% with 402 (1.3%) student responses. Group 4 had 2,353 (7.5%) responses, and Group 5 had the largest number of student responses, 28,436 or 90.4% of the total.

Table 17 shows the Newman-Keuls multiple comparison of means for each category of percentage of class attendance on SATI. The 81-100% (c) group gave the highest mean ratings of 3.1251. Followed by the 61-80% (b) group with a mean rating value of 2.8934. The 41-60%, 0-20%, and 21-40% (a) groupings were statistically indistinguishable from one another, giving mean ratings of 2.7706, 2.7682 and 2.6443, respectively.

Table 17

Mean Comparisons of Faculty Teaching Performance by Percent of Student Classroom Attendance

Groups	Percentage of Attendance	Overall Mean	N
1	0-20% (a) *	2.7682	151
2	21-40% (a)	2.6443	120
3	41-60% (a)	2.7706	402
4	61-80% (b)	2.8934	2353
5	81-100% (c)	3.1251	28436
Overall		3.0997	31,462

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

The relationship between percent of classroom attendance and student mean ratings of faculty teaching performance was positive with the exception of the mean rating of group 1 (0%-20%) being slightly higher (0.1239) than that of group 2 (21-40%), but not statistically significant. The overall mean for all percentage of classroom attendance was

3.0997. Only the 81%-100% attendance percentage grouping was found above the overall mean and then with a nominal value of 0.0254 of a point on a five point scale.

Table 18 shows student percent of class attendance by gender. A total 28,436 (90.5%) students responses reported a class attendance of 81-100% of the time, with a mean faculty teaching performance rating score of 3.1251. The 61-80% level n= 2,353 students (7.4%) out of the total, gave a mean rating of 2.8932. The 41-60% level n= 402 (1.2%), gave a mean of 2.7706, but this mean was statistically indistinguishable from the mean of the 21-40% and 0-20% levels. The percent of the class attended for male and female students was nearly equivalent at all levels of the variable. A Three-way ANOVA was conducted on the variables of percent of class attendance, expected grade, and reported GPA. No interaction was found in the three variable analysis, but main effects were noted for the percent of class attendance variable at p <.001 level of confidence.

Table 18

Student Percent of Class Attendance by Gender

ATTEND		FEMALE F (Count / Row Pct / Col Pct)	MALE M (Count / Row Pct / Col Pct)	Row Total
0-20%	A	62	89	151
		41.1	58.9	.5
		.4	.5	
21-40%	B	40	80	120
		33.1	66.9	.4
		.3	.5	
41-60%	C	167	235	402
		41.5	58.5	1.2
		1.2	1.3	
61-80%	D	1026	1327	2353
		43.6	56.4	7.4
		7.5	7.6	
81-100%	E	12398	16038	28436
		43.6	56.4	90.5
		90.5	90.1	
	Column Total	13693	17769	**31462**
		43.7	56.3	100.0

Chi-Square	Value	DF	Significance
Pearson	6.81886	4	.14578
Likelihood Ratio	6.99323	4	.13625

Minimum Expected Frequency-- 51.591

Hypothesis 6

Hypothesis 6 predicted that there would be no significant difference in how students evaluate faculty across course levels as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results (Table 19) reveal a statistical difference (.05 $F(4,31708) = 2.37 < 13.5491$, $p=.0000$) between course level and student ratings of faculty teaching performance; hence the null hypothesis was rejected.

Table 19

Student Mean Rating Scores of Faculty Teaching Performance by Class Level

Source	df	SS	MS	F
Between groups	4	35.8676	8.9669	13.5491
Within groups	31703	.6618	.6627	
Total	31707	21017.1111		

$p=.0000$

.05 $F(4,31708)=2.37$

A total of 31,708 out of 40,339 student rating responses were included in the class-level (class) variable analysis with 8,631 (21.4%) missing cases or responses failing to indicate class level on SATI. The class-level variable consisted of five levels or groups. Each is listed with the corresponding number of student responses and the percentage that number represents for all levels combined.

The level values are as follows: Group 1 or the 100 level students contained 9,043 (29%) responses; Group 2, contained 7,804 (25%) responses; Group 3, 7,390 (23%) responses; Group 4, 6,091 (19%); and Group 5 with 1,380 (4%) student responses. Group 1 students reported to have attended class between 0-20% of the time for Fall Quarter 1989.

Group 2 ranged from 21-40% with 120 (.38%) of the total number of student responses. The range for Group 3 was 41-60% with 402 (1.3%) student responses. Group 4 had 2,353 (7.47%) responses, and Group 5 had the largest number of student responses, 28,436 or 90.38% of the total.

Table 20

Number and Percentage of Students by Class Level

CLASS STANDING BY COURSE LEVEL

Count / Row Pct / Col Pct		100 (1.00)	200 (2.00)	300 (3.00)	400 (4.00)	500 (5.00)	Row Total
1.00 UNDERGRAD	Count	12000	10378	7487	5190	81	35136
	Row Pct	34.2	29.5	21.3	14.8	.2	91.1
	Col Pct	99.0	98.8	97.9	78.3	4.8	
2.00 GRAD	Count	121	127	164	1437	1599	3448
	Row Pct	3.5	3.7	4.8	41.7	46.4	8.9
	Col Pct	1.0	1.2	2.1	21.7	95.2	
Column Totals		12121	10505	7651	6627	1680	38584
		31.4	27.2	19.8	17.2	4.4	100.0

Table 20 shows the breakdown of the total number and percentage of student rating responses at each course level, and whether they were an undergraduate or graduate student at each level. A total of 12,121 student ratings were noted in the 100 level course, of these 99% were undergraduate and 1% graduate. The 200 level course had 10,505 student ratings, 98.8% were undergraduate students and 1.2% graduate. The 300 level course designate had 7,651 student ratings, 97.9% undergraduate and 2.1 graduate. The 400 level contained 6,627 ratings, 78.3% were undergraduate and 21.7% graduate; and the 500 level course category had 1,680 student ratings of which 4.8% were undergraduate and 95.2% were graduate.

Table 21 shows the results of Newman-Keuls multiple comparison of means for each course level on SATI. The students in the 500 level classes (c) gave the highest mean ratings (3.2034) on faculty teaching performance. Followed by students in the 100 level courses (b) with a mean rating value of 3.1265. The means of the 200, 300, and 400 levels (a) were statistically indistinguishable from one another, giving mean rating values of 3.0677, 3.0692, and 3.0987 respectively.

Table 21

Mean Comparisons of Faculty Teaching Performance by Class Level

Group	Level	Means	N
2	200	3.0677 (a)	7804
3	300	3.0692 (a)	7390
4	400	3.0987 (a)	6091
1	100	3.1265 (b)	9043
5	500	3.2034 (c)	1380
Overall		3.0997	31708

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

The relationship between student mean rating scores of faculty teaching performance and student course level was not linear. The overall mean for all grade categories was 3.0997. Only the means of the 500 level (a) and 100 (b) level courses were above this mean with a mean difference of 0.0268 on a five-point scale.

One-way ANOVA results on student ratings grouped according to graduate and undergraduate standing (Table 22) also indicated that a statistical difference exists (.05 $F(1,38528)=3.84 <92.0313$, $p=.0000$) between graduate and undergraduate students. The graduate group had an overall mean value of 3.2488 points on a five-point scale which was slightly higher than the 3.1235 mean value for the undergraduate group (Table 23).

Table 22

Faculty Teaching Performance in Graduate and Undergraduate Courses

Source	df	SS	MS	F
Between groups	1	58.4150	58.4150	92.0313
Within groups	38526	24565.9490	.6347	
Total	38527	24624.3640		

$p=.0000$

.05 $F(1,38528)=3.84$ 92.0313

Table 23 also depicts student evaluations of faculty teaching performance means at the college level. The differences in the means for graduate and undergraduate students for all colleges was less than .2 on a five-point scale, with the exception of Pharmacy (0.8839), Oceanography (0.3431), and Engineering (0.2685). Graduate student ratings placed the Colleges of Science and Business below the combined college mean, as well as Colleges of Pharmacy, Engineering, and Science by undergraduate students.

Among graduate students the College of Pharmacy received the highest mean ratings, 3.6074, while giving the lowest mean score to the College of Business (3.0107), a mean difference of 0.5967. Conversely, the College of Pharmacy received the lowest mean ratings among colleges by undergraduate students who rated ROTC the highest, with a mean difference of 0.8966 on a five-point scale.

A 3 x 5 x 5 MANOVA was conducted comparing class size, course level, and faculty rank variables with the 12 dependent variables of SATI. Interaction was discovered for the class level variable only on all 12 dependent variables. The highest faculty ratings were given by freshman (3.1265) and graduate (3.2034) student levels. Lowest ratings were given by sophomore (3.0677) and Junior (3.0692) level students. Seniors also showed a low rating value of 3.0987 points on a five-point scale.

Table 23

Faculty Teaching Performance in Graduate and Undergraduate Courses by Colleges.

	Graduate		Undergraduate	
Colleges	Means	N	Means	N
Pharmacy	3.6074	9	2.7235	671
Oceanography	3.5000	1	3.1569	64
Agriculture	3.4760	268	3.3483	1,227
Education	3.4377	382	3.4392	479
Health & PE	3.4019	205	3.4800	3,306
Liberal Arts	3.3992	275	3.1785	10,101
Home Econ	3.3544	141	3.2236	1,251
Engineering	3.2626	434	2.9941	2,952
Science	3.1535	1,271	2.9652	12,014
Business	3.0107	457	3.0578	2,918
ROTC	------	---	3.6201	102
Overall	3.2488	3,443	3.1235	35,085

Hypothesis 7

Hypothesis 7 predicted that there would be no significant difference in how male and female students evaluate faculty across university courses as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results (Table 24) indicate that a statistical difference (.05 $F1,39140=3.84 < 37.2434$, $p=.0000$) was found between male and female student evaluation means of faculty teaching performance; therefore the null hypothesis was rejected.

Table 24

Male and Female Student Ratings of Faculty Teaching Performance

Source	df	SS	MS	F
Between groups	1	23.6910	23.6910	37.2434
Within groups	38138	24594.7074	.6361	
Total	39139	24618.3985		

$p=.0000$

.05 $F(1,39140)=3.84$

A total of 39140 out of 40,339 student rating responses were included in the student gender (student) variable analysis with 1,199 (2.9%) missing cases or responses failing to indicate gender on SATI. Each gender level is listed with the corresponding number of student responses and the percentage that number represents for all levels combined.

The first level represented male student responses of 21,738 (55.6%), and the second contained female students of 17,402 (44.4%) out of a total of 40,339 responses. The overall mean evaluation by male students was 3.1028 and 3.1532 for female students with a mean difference between groups of 0.0504 on a five-point scale (Table 23).

Table 25 provides a Newman-Keuls multiple comparison of student evaluations of faculty teaching performance means across colleges for the male and female student groupings on SATI. The rating means of female and male students were statistically different for most colleges analyzed.

Table 25

Male and Female Student Ratings of Faculty Teaching Performance by Colleges

	Male		Female	
College	Means	N	Means	N
ROTC	3.6258 (a) *	88	3.5317 (a) *	11
Education	3.4878 (b)	339	3.4785 (b)	584
Health	3.4878 (b)	1792	3.4280 (b)	1716
Agriculture	3.3967 (c)	928	3.3424 (c)	596
Home Economics	3.2429 (d)	172	3.2322 (d)	1238
Liberal Arts	3.1621 (e)	4956	3.2020 (d)	5342
Engineering	3.0324 (e)	2934	3.0363 (e)	402
Business	3.0178 (e)	1892	3.0856 (e)	1488
Science	2.9853 (f)	7908	2.9832 (f)	5345
Pharmacy	2.6705 (g)	374	2.7890 (g)	306
Overall	3.1028	21,738	3.1532	17,402
Total				39,140

* Column means followed by the same letter are not statistically different, Newman-Keuls p < .05.

ROTC received the highest mean rating by both male (3.6258) and female (3.5317) students, and the College of Pharmacy the lowest mean ratings by both male (2.6705) and female (2.7890) students, with a difference between means for these academic units of 0.9553 for male students and 0.7427 for female students. The next highest rating scores came from the College of Education with a mean of 3.4878 for male students and 3.4785 for female students. Followed by the Colleges of Health/P.E., Agriculture, Home Economics, Liberal Arts, Engineering, Business, Science, and Pharmacy in descending order mean rating magnitude.

Four groupings were identified in the comparison of teaching performance ratings by male and female students (Table 26): male students rating of male instructors (MSTUD/MINST), female students rating of male instructors (FSTUD/MINST), male students rating of female instructors (MSTUD/FINST), and female student rating of female instructors (FSTUD/FINST). In Figures 1 and 2 significant interaction between gender of the student and gender of the instructor was found and plotted for nine of twelve SATI subscales in the Colleges of Engineering and Home Economics.

In the College of Home economics (Table 26) the range in mean scores for male students rating of male instructors was from a low of 3.3590 on subscale #8 (The instructor encouraged me to think for myself) to a high of 3.6 on subscale #11 (I have learned a significant number of new ideas and skills from the instructor) and subscale #12 (All things considered I was favorably impressed by this instructor). The mean difference between high and low scores was 0.2410.

The low value of female students rating of male instructors was 3.0 on three subscales: 1# (Course objectives/requirements were clearly presented), subscale #9 (Examinations were relevant to assignments and lectures), and subscale #11 (I have learned a significant number of new ideas and skills from the instructor). The high value was 4.0 on subscale #7 (The instructor was fair and impartial in dealing with me), with a rating mean difference between high and low values of 1.0.

The mean scores for male students rating of female instructors ranged from a low of 2.9749 on subscale #3 (The instructor explained the material clearly), to a high rating value of 3.3890 on subscale #7 (The instructor was fair and impartial in dealing with me) with a difference of 0.4780 between high and low rating scores. Female students rating of female instructors ranged from a low of 3.1667 on subscale #11 (I have learned a significant number of new ideas and skills from the instructor), to a high rating value of 3.4609 on subscale #7 (The instructor was fair and impartial in dealing with me). The difference between high and low means was 0.2942.

In the College of Engineering (Table 26) all groups gave the lowest ratings on subscale #3 (The instructor explained the material clearly). In the high score range, two of the four groupings: male students rating of male instructors and female students rating of female instructors selected subscale #2 (The instructor was well prepared and organized) for the highest mean ratings. The remaining two groups, female student rating of male instructors choose subscale #7 (The instructor was fair and impartial in dealing with me) for the high rating score and subscale #11 (I have learned a significant number of new ideas and skills from the instructor), was selected by the male students rating of female instructors group. The differences in high and low mean rating values are as follows: MSTUD/MINST (3.1870 - 2.8696)= 0.3174; FSTUD/MINST (3.2034 - 2.8460)= 0.3574; MSTUD/FINST (3.6000 - 3.3000)= 0.3000, and FSTUD/FINST (3.4231 - 3.0939)= 0.3292.

In comparing mean rating differences between the Colleges of Home Economics and Engineering, subscales that were statistically significant at the .05 level of confidence within the four groups revealed the FSTUD/FINST group with the least variance in mean scores, less than 0.2 on all subscales. The highest difference in favor or Home Economics came from the FSTUD/MINST group with mean rating differences of 0.5 on five of the nine subscales: #3 (The instructor explained the material clearly), subscale #7 (The instructor was fair and impartial in dealing with me), subscale #8 (The instructor encouraged me to think for myself), subscale #10 (The instructor used good communication skills), and #12 (All things considered, I was favorably impressed by this instructor). Of these five subscales just discussed the largest difference (0.8) was found on subscale #3 (The instructor explained the material clearly).

Figure 1. MSMI/FSFI Ratings on Nine Instructional Dimensions of SATI

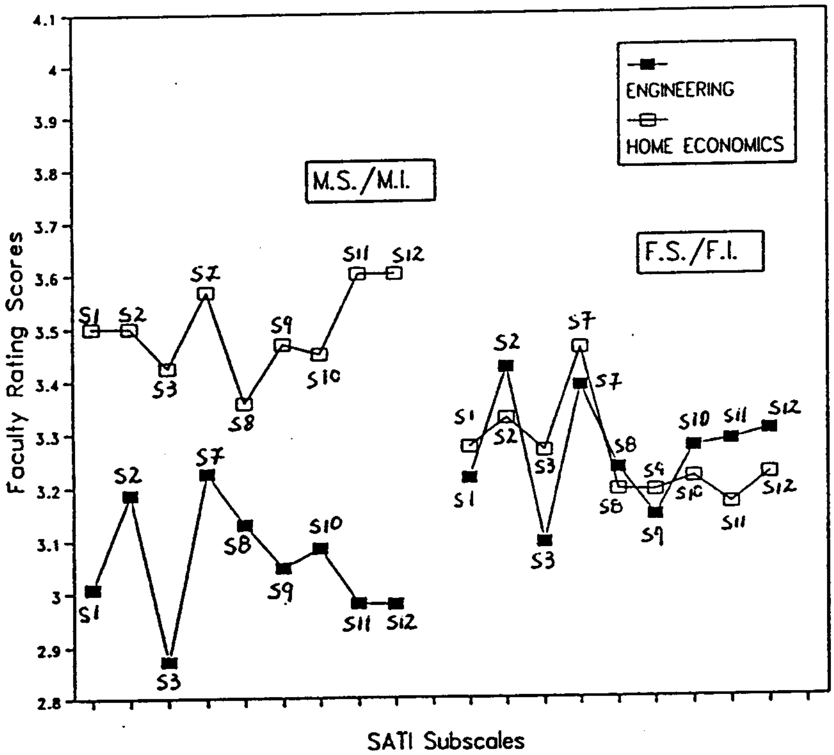


Figure 2. FSMI/MSFI Ratings on Nine Instructional Dimensions of SATI

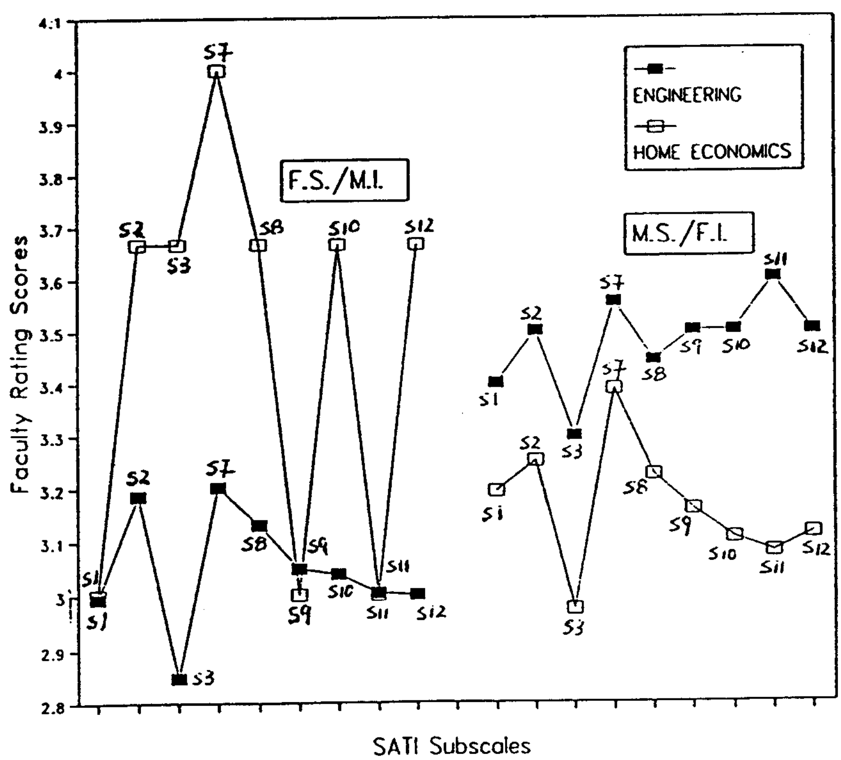


Table 26

Comparisons of Faculty Mean Ratings in the Colleges of Engineering and Home Economics as a Function of Student Gender on SATI

Student	Inst.	Engr.	Home Econ.	Mean Dif.	Subscales
Group 1:					
MSTUD	MINST	3.0082	3.5000	0.4918	1
MSTUD	MINST	3.1870	3.5000	0.3130	2
MSTUD	MINST	2.8696	3.4250	0.5554	3
MSTUD	MINST	3.2265	3.5676	0.3411	7
MSTUD	MINST	3.1304	3.3590	0.2286	8
MSTUD	MINST	3.0476	3.4688	0.4212	9
MSTUD	MINST	3.0845	3.4500	0.3655	10
MSTUD	MINST	2.9808	3.6000	0.6192	11
MSTUD	MINST	2.9783	3.6000	0.6217	12
Group 2:					
FSTUD	MINST	2.9959	3.0000	0.0041	1
FSTUD	MINST	3.1867	3.6667	0.4803	2
FSTUD	MINST	2.8460	3.6667	0.8207	3
FSTUD	MINST	3.2034	4.0000	0.7966	7
FSTUD	MINST	3.1323	3.6667	0.5344	8
FSTUD	MINST	3.0502	3.0000	0.0502	9
FSTUD	MINST	3.0408	3.6667	0.6259	10
FSTUD	MINST	3.0049	3.0000	0.0049	11
FSTUD	MINST	3.0015	3.6667	0.6652	12
Group 3:					
MSTUD	FINST	3.4000	3.1944	0.2056	1
MSTUD	FINST	3.5000	3.2529	0.2471	2
MSTUD	FINST	3.3000	2.9749	0.3251	3

(Table 26 continued)

Student	Inst.	Engr.	Home Econ.	Mean Dif.	Subscales
Group 3:					
MSTUD	FINST	3.5560	3.3890	0.1670	7
MSTUD	FINST	3.4444	3.2265	0.2175	8
MSTUD	FINST	3.5000	3.1603	0.3397	9
MSTUD	FINST	3.5000	3.1072	0.3928	10
MSTUD	FINST	3.6000	3.0799	0.5201	11
MSTUD	FINST	3.5000	3.1145	0.3855	12
Group 4:					
FSTUD	FINST	3.2143	3.2742	0.0599	1
FSTUD	FINST	3.4231	3.3280	0.0951	2
FSTUD	FINST	3.0939	3.2661	0.1722	3
FSTUD	FINST	3.3895	3.4609	0.0714	7
FSTUD	FINST	3.2346	3.1933	0.0413	8
FSTUD	FINST	3.1444	3.1917	0.0473	9
FSTUD	FINST	3.2732	3.2160	0.0572	10
FSTUD	FINST	3.2857	3.1667	0.1190	11
FSTUD	FINST	3.3022	3.2222	0.0800	12

Table 27 contrasts faculty mean ratings by gender of the instructor and student gender and whether the student rating was from a major or non-major course. Two-way ANOVA results for the entire population showed no interaction between student/faculty gender and class standing on 11 of the 12 subscales on SATI. Subscale #7 (The instructor was fair and impartial in dealing with me) showed interaction, Significance of F =.003., at a confidence level of $p < 0.05$. The overall means varied less than 0.1 on a five-point scale between gender of the instructor/student and class status, and less than 0.2 of a point for all subscales except subscale #7, where the mean difference of 0.6 of a point was found between female students' rating of male and female faculty in major courses. A difference of 0.6 of a point in favor of female faculty as rated by female students.

The differences between male and female student mean ratings of faculty teaching performance by college was less than 0.1 of a point, with one exception, Pharmacy had a mean difference of 0.1185 for male and female students. Only the Colleges of Pharmacy, Science, Engineering, and Business were below the overall ratings for all colleges combined.

Table 27

Male and Female Faculty Mean Ratings on 12 Subscales of SATI as a Functionof Student Gender and Major

	Male student (11,034)		Female student (16,186)	
Subscales	Male Faculty (354)	Female Faculty (291)	Male Faculty (354)	Female Faculty (291)
Question 1				
Major	3.2	3.3	3.1	3.1
Non-Major	3.2	3.2	3.2	3.2
Question 2				
Major	3.3	3.4	3.2	3.2
Non-Major	3.3	3.3	3.3	3.3
Question 3				
Major	3.0	3.1	3.0	3.0
Non-Major	3.0	3.1	3.0	3.1
Question 4				
Major	3.0	3.2	3.0	3.1
Non-Major	3.0	3.2	3.1	3.2
Question 5				
Major	3.2	3.3	3.1	3.1
Non-Major	3.0	3.2	3.1	3.1

(Table 27 continued)

Subscales	Male student (11,034) Male Faculty (354)	Male student (11,034) Female Faculty (291)	Female student (16,186) Male Faculty (354)	Female student (16,186) Female Faculty (291)
Question 6				
Major	3.2	3.3	3.1	3.1
Non-Major	3.3	3.4	3.4	3.4
Question 7				
Major	3.4	3.5	3.1	**3.6***
Non-Major	3.3	3.5	3.4	3.4
Question 8				
Major	3.3	3.3	3.2	3.2
Non-Major	3.2	3.2	3.2	3.2
Question 9				
Major	3.2	3.3	3.2	3.1
Non-Major	3.2	3.2	3.2	3.2
Question 10				
Major	3.2	3.2	3.1	3.1
Non-Major	3.1	3.2	3.2	3.2
Question 11				
Major	3.2	3.3	3.1	3.1
Non-Major	3.0	3.1	3.1	3.0
Question 12				
Major	3.2	3.2	3.1	3.1
Non-Major	3.1	3.2	3.2	3.2
Overall				
Major	3.2	3.2	3.1	3.1
Non-Major	3.1	3.3	3.2	3.2

* Statistically significant at the p. < .01. level of confidence.

Hypothesis 8

Hypothesis 8 predicted that there would be no significant difference between course evaluations of male and female faculty as measured by the Student Assessment of Teaching Instrument. Two-way ANOVA results indicate statistical interaction at a probability level of $p < 0.05$ between the variables of gender of the instructor and gender of the student. Table 28 shows interaction discovered on nine of the twelve subscales of SATI for the comparison of gender of the student and gender of the instructor.

Table 28

Mean Comparisons of the Effects of Gender of the Student and Instructor on Instructor Ratings of Teaching Performance

Scale	Student Gender	Instructor Gender	S x I
Question 1			
MS	57.334	21.132	7.365
F	.000**	.000**	.004**
Question 2			
MS	17.744	.192	8.691
F	.000**	----	.000**
Question 3			
MS	.078	29.963	15.192
F	----	.000**	.000**
Question 4			
MS	.265	120.729	2.909
F	----	.000**	----
Question 5			
MS	18.746	27.733	1.984
F	.000**	.000**	----
Question 6			
MS	31.938	18.244	.535
F	.000**	.000**	----
Question 7			
MS	16.973	25.802	5.298
F	.000**	.000**	.009**

(Table 28 continued)

Scale	Student Gender	Instructor Gender	S x I
Question 8			
MS	11.957	.699	13.590
F	.000**	----	.000**
Question 9			
MS	15.804	4.914	7.226
F	.000**	.025*	.007**
Question 10			
MS	.106	8.313	12.191
F	----	.004**	.001**
Question 11			
MS	2.423	1.446	17.016
F	----	----	.000**
Question 12			
MS	1.139	8.186	10.301
F	----	.005**	.002**

Note. df=1 for student sex, instructor sex, and S x I; df= 27,567 for within * $p < .05$. ** $p < .01$.

Table 29 lists those colleges, e.g. Engineering, Home Economics, Education, and Agriculture, with faculty teaching performance means that were significant at .05 level of confidence on student and instructor gender variable. A total of 38,590 out of 40,339 student rating responses were included in the faculty gender variable analysis, with 1,749 (4.3%) missing cases or responses where students failed to indicate the gender of their instructors on SATI. The faculty gender consists of two levels with two groupings per level.

Table 29

Comparison of Male and Female Faculty Mean Ratings by College as a Function of Student Gender

	Male student n = 17,178		Female student n = 21,412	
Colleges	Male Faculty n = 355	Female Faculty n = 290	Male Faculty n = 355	Female Faculty n = 290
Agriculture	3.3*	3.5	3.4	3.2
Education	3.4*	3.5	3.5	3.3
Engineering	3.0*	3.5	3.0	3.3
Home Economics	3.5*	3.2	3.5	3.2
Science	3.0	3.0	3.0	3.0
Liberal Arts	3.2	3.2	3.2	3.2
ROTC	3.5	---	3.6	---
Health/PE	3.4	3.6	3.5	3.5

* College means that were statistically significant at .05 level of confidence.

Table 29 lists each level with the corresponding number of student responses and the percentage that number represents for all levels combined. The first level is male student responses (17,178) with two levels: male faculty (355) and female faculty (290). Level two represents female student responses (21,412) with two levels: male faculty (355) and female faculty (290).

The Colleges of Engineering and Home Economics (Table 29) showed the widest range of mean faculty rating scores and the largest number of interactions on SATI subscales between male and female faculty as rated by male and female students. The difference in means for student and instructor gender in both colleges were significant at the 0.05 level of confidence in the interaction and main effects.

Interaction was found on nine of 12 subscales of SATI: #1 (Course objectives and requirements were clearly presented to me), #2 (The instructor was well prepared and organized), #3 (The instructor explained the material clearly), #7 (The instructor was fair and impartial in dealing with me), #8 (The instructor encouraged me to think for myself), #9 (The examinations were relevant to the reading assignments and the material presented in class), #10 (The instructor used good communications skills), #11 (As a result of having this instructor, I have learned a significant number of new ideas and/skills), and #12 (All things considered, I was favorably impressed by this instructor). Significant main effects were found on subscales #4 (The instructor was sensitive to my/the class' ability to understand the material), #5, (The instructor stimulated enthusiasm for the subject matter of the course), and #6 (The instructor provided scheduled office hours or was readily available for consultation with me).

The Colleges of Science and Liberal Arts showed no interaction and had the narrowest range of overall mean rating scores of all colleges on SATI subscales between male and female instructors as a function of student gender (Table 30). There were no female instructors from ROTC represented in this study.

Table 30

Male and Female Faculty Mean Ratings on SATI Subscales by College as a Function of Student Gender

	Male student n = 17,178		Female student n = 21,412	
Colleges	Male Faculty n = 355	Female Faculty n = 290	Male Faculty n = 355	Female Faculty n = 290
Question 1				
Agriculture	3.3	3.6	3.5	3.2
Education	3.4	3.5	3.5	3.3
Engineering	3.0	3.4	3.0	3.2
Home Economics	3.5	3.1	3.2	3.0
Science	3.1	3.1	3.0	3.1
Liberal Arts	3.2	3.3	3.2	3.1
ROTC	3.6	---	3.7	---
Health/PE	3.5	3.6	3.5	3.6
Question 2				
Agriculture	3.3	3.7	3.5	3.1
Education	3.3	3.6	3.5	3.5
Engineering	3.1	3.5	3.2	3.4
Home Economics	3.6	3.2	3.7	3.3
Science	3.2	3.2	3.2	3.2
Liberal Arts	3.3	3.3	3.2	3.2
ROTC	3.4	---	3.8	---
Health/PE	3.6	3.6	3.6	3.6
Question 3				
Agriculture	3.1	3.4	3.3	3.2

(Table 30 continued)

	Male student n = 17,178		Female student n = 21,412	
Colleges	Male Faculty n = 355	Female Faculty n = 290	Male Faculty n = 355	Female Faculty n = 290
Question 3				
Education	3.2	3.3	3.4	3.2
Engineering	2.9	3.3	2.8	3.1
Home Economics	3.4	2.9	3.7	3.3
Science	2.8	2.9	2.9	2.9
Liberal Arts	3.0	3.1	3.0	3.0
ROTC	3.5	---	3.5	---
Health/PE	3.3	3.5	3.4	3.5
Question 4				
Agriculture	3.2	3.5	3.4	3.3
Education	3.4	3.4	3.5	3.4
Engineering	2.9	3.7	2.9	3.3
Home Economics	3.6	3.1	3.7	3.1
Science	2.9	3.0	2.9	3.0
Liberal Arts	3.0	3.2	3.0	3.1
ROTC	3.4	---	3.5	---
Health/PE	3.4	3.6	3.4	3.5
Question 5				
Agriculture	3.3	3.5	3.7	3.2
Education	3.4	3.4	3.5	3.2
Engineering	2.8	3.6	2.8	3.1
Home Economics	3.6	3.0	3.7	3.2

(Table 30 continued)

Colleges	Male student n = 17,178 Male Faculty n = 355	Male student n = 17,178 Female Faculty n = 290	Female student n = 21,412 Male Faculty n = 355	Female student n = 21,412 Female Faculty n = 290
Question 5				
Science	2.9	2.8	2.9	2.9
Liberal Arts	3.1	3.2	3.1	3.1
ROTC	3.5	---	3.7	---
Health/PE	3.4	3.6	3.4	3.5
Question 6				
Agriculture	3.4	3.4	3.5	3.2
Education	3.4	3.4	3.5	3.2
Engineering	3.1	3.3	3.2	3.3
Home Economics	3.4	3.4	3.0	3.3
Science	3.3	3.3	3.2	3.3
Liberal Arts	3.4	3.5	3.4	3.4
ROTC	3.5	---	3.8	---
Health/PE	3.5	3.4	3.4	3.4
Question 7				
Agriculture	3.5	3.7	3.5	3.3
Education	3.6	3.6	3.6	3.5
Engineering	3.2	3.6	3.2	3.4
Home Economics	3.6	3.3	4.0	3.5
Science	3.3	3.4	3.3	3.4
Liberal Arts	3.3	3.4	3.3	3.3
ROTC	3.8	---	3.8	---

(Table 30 continued)

	Male student n = 17,178		Female student n = 21,412	
Colleges	Male Faculty n = 355	Female Faculty n = 290	Male Faculty n = 355	Female Faculty n = 290
Question 7				
Health/PE	3.6	3.7	3.6	3.6
Question 8				
Agriculture	3.9	3.1	3.4	3.2
Education	3.6	3.5	3.7	3.3
Engineering	3.1	3.4	3.1	3.2
Home Economics	3.4	3.2	3.7	3.1
Science	3.1	3.1	3.1	3.1
Liberal Arts	3.3	3.4	3.3	3.3
ROTC	3.7	---	3.4	---
Health/PE	3.4	3.5	3.4	3.4
Question 9				
Agriculture	3.4	3.7	3.5	3.3
Education	3.6	3.6	3.7	3.4
Engineering	3.0	3.5	3.0	3.1
Home Economics	3.5	3.2	3.7	3.1
Science	3.0	3.2	3.1	3.2
Liberal Arts	3.3	3.3	3.2	3.1
ROTC	3.6	---	3.7	---
Health/PE	3.5	3.6	3.5	3.5
Question 10				
Agriculture	3.3	3.5	3.4	3.2

(Table 30 continued)

	Male student n = 17,178		Female student n = 21,412	
Colleges	Male Faculty n = 355	Female Faculty n = 290	Male Faculty n = 355	Female Faculty n = 290
Question 10				
Education	3.4	3.5	3.6	3.4
Engineering	3.1	3.5	3.0	3.3
Home Economics	3.5	3.1	3.7	3.2
Science	3.0	3.0	3.0	3.0
Liberal Arts	3.1	3.2	3.2	3.2
ROTC	3.4	---	3.6	---
Health/PE	3.4	3.5	3.5	3.4
Question 11				
Agriculture	3.3	3.4	3.5	3.3
Education	3.4	3.4	3.5	3.3
Engineering	2.9	3.6	3.0	3.2
Home Economics	3.6	3.0	3.0	3.2
Science	3.0	3.0	3.0	3.0
Liberal Arts	3.0	3.1	3.0	3.0
ROTC	3.5	---	3.4	---
Health/PE	3.4	3.5	3.4	3.5
Question 12				
Agriculture	3.4	3.4	3.6	3.2
Education	3.4	3.4	3.6	3.2
Engineering	3.0	3.5	3.0	3.3
Home Economics	3.6	3.1	3.7	3.2
Science	3.0	3.0	3.0	3.0
Liberal Arts	3.1	3.2	3.1	3.1
ROTC	3.4	---	3.7	---
Health/PE	3.4	3.6	3.5	3.6

Hypothesis 9

Hypothesis 9 predicted that there would be no significant difference in class size on how students evaluate faculty across all courses as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results (Table 31) indicate no statistical difference (.05 F1,1733=3.84 > 2.0391, p=.1535) between the variables of class size and student evaluation of faculty teaching performance; therefore the null hypothesis was retained.

Table 31

Effects of Class Size on Student Ratings of Faculty Teaching Performance (Based on Minimum Class Size Policy)

Source	df	SS	MS	F
Between groups	1	.4702	.4702	2.0391
Within groups	1731	399.1586	.2306	
Total	1732	399.6288		

p=.1535

.05 F(1,1733)=3.84

The unit of analysis was the class in this hypothesis. Class size was based on the Minimum Class Size Policy set by the Registrars Office at Oregon State University. The minimum number of students allowed per undergraduate and graduate classes are as follows: lower division = 16, upper division = 12, and graduate = 8 students per class. Student numbers below these bench marks represent small classes and above these bench marks represent large classes.

The number of classes in this analysis was 780 for the small class size and 953 for the large class size with a total of 1733 classes in the study. The overall mean was 23.3 and median 18.0 for the number of students per class. Table 32 lists the mean rating scores of faculty teaching performance by college for small and large class sizes. The overall mean for the small group was 4.0445 and 4.0124 for the large group with a difference of 0.0321 between large and small groups.

Table 32

Effects of Class Size on Student Ratings of Faculty Teaching Performance by College (Based on the Minimum Class Size Policy)

	Small		Large	
Colleges	Means	N	Means	N
Health & PE	4.2872	66	4.2317	120
Home Economics	4.2816	26	4.1804	31
Education	4.2623	46	4.2430	23
Agriculture	4.2517	62	4.1977	35
Business	4.2004	11	4.0319	96
Engineering	4.1236	46	3.9390	79
Liberal Arts	4.0830	192	4.0797	285
Pharmacy	4.0578	2	3.6277	10
Oceanography	4.0431	3	4.0575	1
Science	3.8705	326	3.8087	270
ROTC	------	--	4.4985	3
Overall	4.0445	780	4.0124	953

The College of Health/P.E. received the highest mean rating of 4.2872 for the small group, whereas ROTC had the highest mean ratings of 4.4985 in the large group. There was a mean difference between high ratings in the small and large groups of <u>0.21</u> of a point on a five point scale. The lowest mean ratings came from the College of Science (3.8705) in the small group designate, and from the College of Pharmacy with a mean of 3.6277 for the large group. There was a mean difference between low ratings in the small and large groups of <u>0.04</u> of a point on a five point scale.

The difference between mean ratings of faculty teaching performance in small and large groups was less than 0.2 of a point on a five point scale within all colleges and between all colleges with one exception, the College of Pharmacy had a mean difference of <u>0.4301</u>.

Further analysis was conducted on the class size variable and its effect on faculty ratings. Classes were regrouped in two ways: 1. according to natural divisions in the data set, and 2. according to a number above or below 12 students per class. A significant difference (.05 $F_{3,32133} = 2.60 < 136.3870$, $p=.0219$) was found when class sizes were determined according to natural divisions in the data set (Table 33). A statistically significant difference (.05 $F_{1,1733} = 3.84 < 5.2640$, $p=.0219$) was found when class size was grouped above or below 12 students per class (Table 34).

Table 33

Effects of Class Size on Student Ratings of Faculty Teaching Performance (Based on Previous Research Utilizing the 1-28, 29-50, 51-100, and 100+ Groupings)

Source	df	SS	MS	F
Between groups	3	267.8471	89.2824	136.3870
Within groups	32129	21032.4495	.6546	
Total	32132	21300.2966		

p=.0000

.05 F(3,32133)=2.60

Table 34

Effects of Class Size on Student Ratings of Faculty Teaching Performance (Based on Previous Research Utilizing Above and Below the 12 Students Per Class)

Source	df	SS	MS	F
Between groups	1	1.2116	1.2116	5.2640
Within groups	1731	398.4172	.2302	
Total	1732	399.6288		

p=..0219

.05 F(1,1733)=3.84   SMALL=786 classes/LARGE=947=TOTAL OF 1733

The first regrouping was conducted on the class size variable based on natural divisions in the data set. Figure 3 shows the distribution of the number of classes and students in each range category or natural division. The 13-24 students per class range constituted the largest segment of classes representing 770 classes or 44.43% of the total number of classes in the study.

Figure 3

Histogram of Class Size Distribution For All Classes in the Study

```
Classes  Range   * =16 classes

424        0-12  I***************************
770       13-24  I************************************************
282       25-36  I******************
107       37-48  I*******
50        49-60  I***
36        61-72  I**
17        73-84  I*
12        85-96  I*
11       97-108  I*
10      109-120  I*
5       121-132  I
2       133-144  I
1       145-156  I
3       157-168  I
1       169-180  I
0       181-192  I
1       193-205  I
0       206-217  I
0       218-229  I
0       230-241  I
1       242-253  I
                +----+----+----+----+----+----+----+----+----+----+
                0         160       320       480       640       800

                               Histogram Frequency

Mean= 23.3     Median= 18.0    Mode= 14.0   Min.= 1.0    Max.= 251.0
Valid cases= 1733              Missing cases= 0
```

The next highest was the 0-12 range with 424 classes or 24.46% of the total number of classes, followed by the 25-36 range with 282 classes (16.27%); the 37-48 range with 107 classes (6.17%); 49-60 range, 48 classes (2.77%); and the 61-72 range with 32 classes (1.85%). The 73-84, 85-96, 97-108, and 109-120 ranges had 16 classes each or 0.92% each, or collectively 3.68% of the total number of classes. The ranges with the largest number of students in the fewest classes were represented by the 169-180, 193-205, and 242-253 ranges with 1 class (0.057%) each, or collectively 0.17% of the total number of classes.

Table 35 shows the Newman-Keuls multiple comparison of means for each class size category gathered by SATI. The 1-28 (a) group gave the highest mean ratings of 3.1866, but statistically indistinguishable from the mean of 3.1152 from the 29-50 grouping. The 51-100 and 100+ (b) groupings were also statistically indistinguishable from each other, giving mean ratings of 2.9743 and 2.9811, respectively. The difference between student mean ratings based on class size was small (0.2123). Only the means of groups 1 and 2 were found above the overall mean and then with a nominal value of 0.01 of a point on a five-point scale.

The second regrouping of the class size variable data set was conducted based on the number of students per class above and below 12 student per class. The second regrouping consisted of two groupings, A and B. Group A contained 491 (28.33%) classes ranging from 1 to 12 students per class, and B with 1242 (71.67%) classes 13 or more students per class. Group A had a mean of 4.0707 and B a mean of 4.0121, and a difference between mean A and mean B of 0.059 on a five-point scale.

Table 35

Mean Comparisons of Faculty Teaching Performance by Class Size and Number of Classes

Groups	Size	Means	N
1	1-28	3.1866 (a) *	1,328
2	29-50	3.1152 (a)	288
3	51-100	2.9743 (b)	101
4	100+	2.9811 (b)	16
Overall		3.0987	1733

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Hypothesis 10

Hypothesis 10 predicted that there would be no significant difference in course evaluations between tenured and nontenure faculty as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results (Table 34) indicated no statistically significant difference (.05 $F_{1,64}$=3.84 > 1.0050, p= 0.3165) between tenured and non-tenured faculty; therefore the null hypothesis was retained.

Table 36

Student Ratings of Tenured and Nontenure Faculty Teaching Performance

Source	df	SS	MS	F
Between groups	1	.2019	.2019	1.0050
Within groups	643	129.1728	.2009	
Total	644	129.3747		

p=.3165

.05 F(1,645)=3.84

A total of 645 faculty members were included in the tenure/non-tenure (faculty) variable analysis. Tenured faculty constituted 336 (52%) of the total, and the remaining 309 (48%) for the non-tenured group. The overall mean evaluation for tenured faculty was 4.0351, and 4.0705 for non-tenured faculty, with a difference between tenure status of 0.04 on a five-point scale.

Table 37 presents a Newman-Keuls multiple comparison of three tenure status groups, the means of which are statistically indistinguishable from one another, and Table 39 with the number and percentage of faculty in each group: fixed, n = 154 (24%), annual, n = 155 (24%), and indefinite, n = 336 (52%) of the total number of faculty in the study.

Table 37

Mean Comparisons of Student Ratings of Faculty Teaching Performance by Tenure Status

Tenure	Means	N
Fixed	4.0593 (a)*	154
Annual	4.0816 (a)	155
Indefinite	4.0351 (a)	336
Totals	4.0520	645

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Table 38 provides a Newman-Keuls multiple comparison of student evaluations of faculty teaching performance means across colleges for the tenured and non-tenured groups. A tenure status demographic variable does not appear on SATI. The ROTC (a) received the highest mean ratings of all colleges, for both tenured (4.5976) and non-tenure (4.4489) groups.

The next highest mean ratings in order of magnitude for the tenured group were those in the Colleges of Education (4.3745), Home Economics (4.2824) Agriculture (4.2456), Health & P.E., and Liberal Arts (b), which were grouped together as being statistically the same. The lowest mean ratings were received by the College of Pharmacy (3.7406) (d). The Colleges of Business (3.9817), Science (3.9284), Engineering (3.9246) and Pharmacy (3.7406) were all below overall mean for all colleges combined.

Pearson product moment correlation coefficients were computed for each tenure level. Results yielded a correlation of .0221; for all tenure levels combined, corresponding to $p= .05$ value; for the fixed term tenure status r= .0373, annual .0296, and indefinite .0162. Low positive or negative correlations indicate that no systematic relation-ship exists between faculty tenure status and mean ratings of teaching performance for all tenure levels.

Table 38

Student Ratings of Tenured and Nontenure Faculty Teaching Performance by Colleges

	Tenured		Nontenure	
Colleges	Means	N	Means	N
ROTC	4.5976 (a) *	1	4.4489	2
Education	4.3745 (b)	10	3.9799	19
Home Econ	4.2824 (b)	13	3.9716	17
Agriculture	4.2456 (b)	34	4.2177	23
Health & PE	4.1283 (b)	13	4.2139	27
Liberal Arts	4.0810 (b)	83	4.0741	110
Business	3.9817 (c)	24	4.0639	28
Science	3.9284 (c)	102	3.9473	49
Engineering	3.9246 (c)	47	4.1062	34
Pharmacy	3.7406 (d)	8	------	---
Overall	4.0351	336	4.0705	309

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Table 39 shows the number and percentage of faculty in each academic rank and tenure status. The instructor rank has 128 members, six in the indefinite tenure states with the largest percentage in the fixed term designate. Assistant professors number 177 with 28 in fixed term, 140 in annual track, and nine in the indefinite tenure level. Associate professors have a total of 163 members: four fixed term, 15 annual track, and 144 at the indefinite tenure level. Full professors numbered 177 with all represented in the indefinite tenure bracket.

Table 39

Tenure Status of All Academic Ranks by Total Numbers and Percentages of Representation in the Population

RANK	Count / Row Pct / Col Pct	FIXED RANK 1.00	ANNUAL TENURE TRACK 2.00	INDEFINITE TENURE 3.00	Row Total
1.00 INSTRUCTOR	Count	122		6	128
	Row Pct	95.3		4.7	19.9
	Col Pct	28.8		1.8	
2.00 ASSIST PROF	Count	28	140	9	177
	Row Pct	15.8	79.1	5.1	27.4
	Col Pct	6.6	89.2	2.7	
3.00 ASSOC PROF	Count	4	15	144	163
	Row Pct	2.5	9.2	88.3	25.3
	Col Pct	9	9.6	42.7	
4.00 PROFESSOR	Count			177	177
	Row Pct			100.0	27.4
	Col Pct			52.5	
Column Total		154	155	336	**645**
		23.9	24.0	52.1	100.0

Hypothesis 11

Hypothesis 11 predicted that there would be no significant difference in the course evaluations of non-international graduate teaching assistants and full-time faculty members as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results in Table 40 showed no statistically significant difference (.05 F6,887=2.21 < 8.3778, p=.0000) between non-international graduate teaching assistants (NITAs) and full-time faculty (F-TF) members; therefore, the null hypothesis was rejected.

Table 40

Teaching Performance Ratings of Graduate Teaching Assistants and Full-Time Faculty

Source	df	SS	MS	F
Between groups	6	8.1004	1.6201	8.3778
Within groups	881	170.3647	.1934	
Total	886	178.4651		

p=.0000

.05 F(6,887)=2.21

A total of 193 NITAs and 645 F-FT members were included in faculty/ T.A. (faculty) variable analysis. The faculty level variable consisted of four levels. Each is listed with the corresponding number of faculty and the percentage that number represents for all levels combined: instructors, n = 128 (20%), assistant professors, n = 177 (27%) associate professors, n = 163 (26%), and full professors, n = 177 (27%), Table 38.

The overall mean evaluation for NITA was 3.9866, and 4.0345 for F-TF members, with a difference between groups of 0.05 on a five-point scale.

Table 41 provides a Newman-Keuls multiple comparison of student evaluations of faculty teaching performance means by academic rank for NITAs and F-TF member categories.

Table 41

Mean Comparisons of the Teaching Performance of Graduate Teaching Assistants and Full-Time Faculty

Academic Ranks	Overall Means	N
1 Full Professors	4.0041 (a) *	177
2 Associate Professors	4.0562 (a)	163
3 Assistant Professors	4.0826 (a)	177
4 Instructors	4.0709 (a)	128
5 Non-international T.A.s	3.9866 (a)	193
6 International T.A.s	3.6525 (b)	49
Totals	4.0157	887

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

The results of this comparison point to mean rating values that are statistically indistinguishable from one another. Table 42 provides teaching performance ratings of NITA as compared to F-TF academic ranks by colleges. Faculty ratings were not statistically different at all academic ranks from the NITAs; however ITA mean ratings were significantly different from other academic ranks. At the college level of analysis, mean ratings for ITAs in the Colleges of Business, Health/P.E., and Liberal arts were less than 4.0 p on a five-point scale. The Colleges of Science and Education received the lowest overall mean rating, 3.5069 and 3.3194 respectively.

Table 42

Teaching Performance Ratings of Graduate Teaching Assistants and Full-Time Faculty by Academic Rank and Colleges

	Faculty Means				T.A. Means	
Colleges	Inst.	Asst.	Assoc.	Prof.	NITA	ITA
Pharmacy	------	------	3.4688	4.0124	------	------
Science	4.0175	3.9287	3.9959	3.8846	3.8922	3.5069
Agriculture	4.4825	4.1619	4.2481	4.2438	4.0238	------
Engineering	4.0795	4.1160	3.9275	3.9624	------	------
Business	4.0546	4.0844	4.0958	3.9069	3.9172	4.1087
Home Econ.	4.0691	4.0538	4.1136	4.3834	4.1927	3.7564
Health/P.E.	4.2280	4.1196	4.2721	3.8690	4.4279	4.1692
Lib. Arts	4.0811	4.0941	4.0826	4.0615	4.0811	4.2089
Education	3.5086	4.1707	4.2412	4.5810	3.8362	3.3194
Total: Means	4.0709	4.0826	4.0562	4.0041	3.9866	3.6525
Total: n =	128	177	163	177	193	49

Hypothesis 12

Hypothesis 12 predicted that there would be no significant difference in course evaluations between non-international and international graduated teaching assistants as measured by the Student Assessment of Teaching Instrument. One-way ANOVA results in Table 43 indicate that a statistically significant difference (.05 F,242=3.84 < 25.1590, p=.0000) was found between non-international graduate teaching assistants (NITA) and international teaching assistants (ITA); hence the null hypothesis was rejected.

Table 43

Student Ratings of International and Non-International Teaching Assistants

Source	df	SS	MS	F
Between groups	1	4.3620	4.3620	25.1590
Within groups	240	41.6109	.1734	
Total	241	45.9729		

p=.0000

.05 F(242)=3.84

A total of 193 NITA's and 49 ITA's were included in the T.A. variable analysis. The ITA group had an overall mean value of 3.6525 on a five-point scale compared to a value of 3.9866 for the NITA group. The difference in mean evaluation scores was .3341 on a five-point scale (Table 41). Table 44 shows the results of a two-way ANOVA comparing the mean ratings of ITA and NITA groups with student class status (freshman, sophomore, junior and senior) on five of the twelve subscales on SATI. Subscale #1, #3, #9, #10, and #11 showed interaction between the class status and TA variables at a confidence level of .05.

Table 44

Comparisons of Mean Ratings of Non-International and International Teaching Assistants with Class Status on SATI

Class Status	NITA	ITA	Mean Dif.	Subscales
FRESHMAN	4.1372	3.8766	0.2606	1
	3.9798	3.5890	0.3908	3
	3.4779	3.3981	0.0798	9
	4.1082	3.3881	0.7201	10
	3.8237	3.4653	0.3584	11
SOPHOMORE	4.1558	3.7862	0.3696	1
	4.1741	3.5370	0.6371	3
	3.7957	3.7705	0.0252	9
	4.1849	3.1737	1.0112	10
	4.0308	3.0400	0.4804	11
JUNIOR	4.0394	3.7532	0.2862	1
	4.0970	3.9018	0.1952	3
	3.8054	2.9565	0.8489	9
	4.1560	3.6116	0.5444	10
	4.1139	3.9797	0.1342	11
SENIOR	3.7428	2.6269	1.1159	1
	4.0322	3.0365	0.9957	3
	2.2938	1.2487	0.4212	9
	4.0654	3.2452	0.8202	10
	3.9330	2.8827	1.0503	11

(Table 44 continued)

Overall Mean Differences

	FRESHMAN	SOPHOMORE	JUNIOR	SENIOR
NITA	0.6593	0.3892	0.3085	1.7384
ITA	0.4885	0.6125	1.0232	1.9965

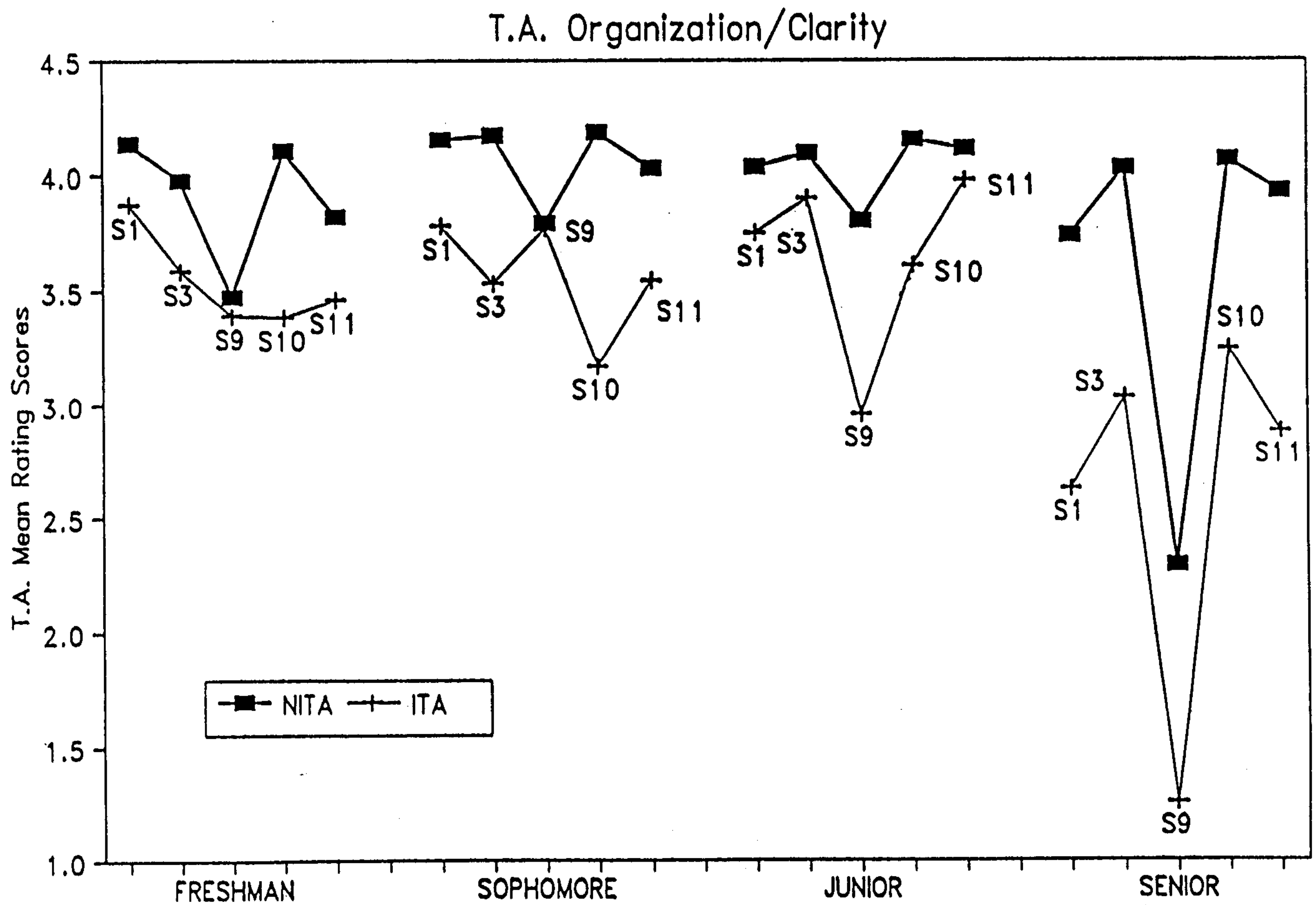
The freshman in the ITA group had the smallest range in mean rating scores on all SATI subscales combined from a high (3.8766) on subscale #1 (Course objectives/requirements were clearly presented), to a low (3.3881) on subscale #10 (The instructor used good communications skills), a point difference of 0.4885. Sophomores also gave a mean rating high on subscale #1 of 3.7862 and a low on subscale #11 of 3.0400, a difference of 0.7462.

Juniors gave a rating high of 3.9797 on subscale #11 (I have learned a significant number of new ideas and skills from the instructor) and a low of 2.9565 on subscale #9 (Examinations were relevant to assignments and lecture), with a difference of 1.0232. Seniors gave a mean rating high of 3.2452 on subscale #10 and a low value of 1.2487 for subscale #9 with a difference of 1.9965.

In the NITA group, the low mean rating values for all class levels were given on subscale #9 (Examinations were relevant to assignments and lecture) with a corresponding high rating value given on subscale #10 (The instructor used good communication skills); with one exception, freshman gave the high ratings for subscale #11 (I have learned a significant number of new ideas and skill from the instructor), as did the freshman in the ITA group. The difference in low-high scores for each class level in the NITA group are as follows: freshman 0.6593, sophomores 0.3892, juniors 0.3506, and seniors 1.7716. Figure 4 shows interaction results on five SATI subscales for all student class levels.

Figure 4

Comparison of Teaching Performance Ratings of ITAs and NITAs on Five Instructional Dimensions of SATI



On subscale #10 alone (The instructor used good communication skills) the difference between ITA and NITA groups for all course levels was greater than .5 on a five-point scale, and greater than 1.00 for sophomores. The ITA and NITA groups also differed by 1.00 point on subscale #1 (Course objectives and requirements were clearly presented) and subscale #9 (Examinations were relevant to assignments and lecture) for the senior class level. On subscale #11 (I have learned a significant number of new ideas and skill from the instructor) seniors received a mean rating greater than.

The ITA group was subsequently divided into categories on the bases of individual participation in either the ITA Orientation Seminar in Postsecondary and Technological Education, participation in teacher preparation at the department level, both seminar and department levels, or on the basis of no formal training at O.S.U. International teaching assistants totaled 49 and were divided into four groups:

1. TAs who had formal pedagogical training at the departmental level before being assigned teaching responsibilities,
2. TAs who had formal pedagogical preparation as a TA through the ITA Orientation Seminar offered by Postsecondary and Technological Education,
3. TAs who had pedagogical training at the department level and ITA orientation program at least once, and
4. TAs who had no formal pedagogical preparation as a TA before teaching Fall term.

Teaching performance means for each TA was computed separately; then a composite score obtained for the group in which TAs were assigned for the analysis. Table 45 shows the Newman-Keuls multiple comparison of means for each TA training category gathered by SATI. The departmental level (group 1) received the highest mean ratings of 3.8863, which was statistically higher than the means of all other groups. This mean was also significantly higher than the mean of 3.6525 for all ITAs combined.

Table 45

Mean Comparisons of the Teaching Performance of Graduate Teaching Assistants in TA Training Programs

Group	Overall Means	N
1 Departmental	3.8863 (a) *	31
2 ITA Orientation Seminar	3.4866 (b)	15
3 Both 1 and 2	3.4817 (b)	28
4 No Prior O.S.U. Training	3.2976 (c)	3
Totals	3.6525	49

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

The second highest mean was 3.4866 received by the ITA Orientation Seminar (group 2), which was statistically indistinguishable from the mean of 3.4817 received by TAs who attended both types of training programs (group 3) training programs. The lowest mean ratings were given to the ITAs in Group 4 who did not attend any formal teacher training program. The mean for this group was substantially below the overall mean for all groups combined.

Hypothesis 13

Hypothesis 13 predicted that there would be no significant difference between course evaluations and the publication rates among full-time faculty members by academic rank and college. There were 645 full-time faculty in the Pearson product moment correlation analysis of ratings with publication rates by academic ranks.

Academic ranks consisted of four levels, which are listed with the corresponding number of faculty and the percentage that number represents for all levels combined: instructors, n = 128 (20%), assistant professors, N = 177 (28%) associate professors, n = 163 (25%), and full professors, n = 176 (27%). Faculty publication rate data were averaged for each academic rank. Mean publication rates were then correlated with teaching performance means for each academic rank and for all academic ranks combined (Table 46).

Table 46

Correlation Coefficients Comparing Full-Time Faculty Teaching Performance Ratings and Faculty Publication Rates by Academic Rank

Faculty Rank	Cor. Coef.	Means	Pub. Rates	N
Instructor	.0487	4.0709	1.0763 (a)*	128
Assistant Prof.	-.0248	4.0826	4.9450 (b)	177
Associate Prof.	-.0297	4.0562	6.1290 (c)	163
Full Prof.	.0099	4.0041	5.9604 (c)	177
Overall	-.0293	4.0134	4.8879	645

* Column means followed by the same letter are not statistically different, Newman-Keuls $p < .05$.

Pearson product moment correlation results yielded a coefficient value of -.0293 for faculty ranks overall, which corresponds to a p=.05 value. The negative relationship between course evaluations and publication rates overall was also true for assistant professors (-.0248) and associate professors (-.0297). Instructors had a positive correlation of .0487 as did full professors (.0099). Low overall coefficients indicates that no systematic relationship exists between faculty publication rates and their mean rating scores of teaching performance for all academic ranks.

Table 47 lists correlation coefficients of all academic ranks and Table 48 by college. The following colleges had positive correlation values indicating no systematic relationship exists between faculty publication rates and their mean rating scores of teaching performance: Agriculture (.0819), Business (.0896), Health/PE (.0015), and Science (.0717). Colleges with low positive correlations were Education (.2015) and Engineering (.1026).

The College of Home Economics had a moderately positive correlation value of .4488, while the College of Pharmacy had a moderately negative value of (-.3143). Only the College of Liberal Arts had a high correlation value, and it was a -0.9635.

Faculty publication rate data were collected for 644 full- time faculty over a five year period, the results are also provided in Table 48 by academic rank and college. The Colleges with the highest overall publication rates were Pharmacy (13.0), Science (11.6513), and Agriculture (10.0702). Colleges with the lowest means were Health/PE (2.2750), Liberal Arts (1.9635), and Education (.5868).

The range in publication rates between colleges was from a low of 0.5862 in the College of Education to a high of 13.0 in the College of Pharmacy, with a difference of 12.4138 publications over a five-year period between the two colleges. The mean publication rate for all colleges combined was 4.8879 articles published over five years, or approximately one article per year at the college and academic rank levels. Instructors and assistant professors were not represented in the data for the College of Pharmacy.

Table 47

Correlation Coefficients of Full-Time Faculty Teaching Performance Ratings and Faculty Publication Rates by Colleges

College	Cor. Coef.	SATI Means	Pub. Rates	N
Agriculture	.0819	3.3676	10.0702	57
Business	.0896	3.0492	2.7115	52
Education	.2015	3.4462	.5862	29
Engineering	.1026	3.0250	7.8519	81
Health & PE	.0015	3.4762	2.1750	40
Home Econ.	.4488*	3.2351	2.6333	30
Liberal Arts	-.9635	3.1817	1.9635	192
Pharmacy	-.3143	2.7325	13.0000	8
ROTC	-----	3.6201	------	3
Science	.0717	2.9835	11.6513	152
Oceanography	.1261	3.1764	7.9625	1
Overall	-.0291	3.1229	4.8879	645

Table 48

Faculty Publication Rates in Descending Order by College and Academic Rank at Oregon State University

	Publication Means					
Colleges	Inst.	Asst.	Assoc.	Prof.	Overall	N
Pharmacy	------	-------	15.2500	7.8475	13.0000	8
Science	2.1667	11.4194	10.8824	14.8088	11.6513	152
Agriculture	1.7500	9.4737	7.9286	13.0053	10.0702	57
Engineering	3.2857	6.4091	10.0645	7.6190	7.8519	81
Oceanography	-----	------	-----	7.0430	7.0430	1
Business	0.5000	2.7500	4.7778	3.8000	2.7115	52
Home Econ.	0.3333	1.5333	5.0000	3.3333	2.6333	30
Health & P.E.	0.4375	5.0000	2.2727	1.6667	2.1750	40
Liberal Arts	0.1379	2.4038	2.700	3.2368	1.9635	192
Education	0.0000	0.5714	1.2857	0.0000	0.5862	29
Totals: Means	1.0763	4.9450	6.1290	5.9604	4.8879	645
Total: n =	128	177	163	177		645

Instructors (Table 48) publish least of all academic ranks with an overall mean of 1.0763 publications in five years. Associate professors publish the most with a mean value of 6.1290 for all colleges combined. The range between the high and low publication values for academic rank was 5.0527. Assistant professors had an overall mean publication rate of 4.9450; however, in the Colleges of Science, Engineering, Agriculture, and Health/P.E. their publication averages exceeded that noted for all assistant professors combined. Conversely, the Colleges of Home Economics, Liberal Arts, Business, and Education were below the average at the assistant professor rank. This was also true for the associate professor rank with Pharmacy replacing Health/PE in the first grouping of colleges.

Full professors showed an overall publication mean of 5.9604 articles published in five years. In the Colleges of Science, Agriculture, and Liberal Arts the full professor publication means exceeded that of all other academic ranks.

Table 49 provides a contrast of faculty publication rate data from Oregon State University College of Education with data collected from five randomly selected Colleges of Education within universities belonging to the Pacific Athletic Conference (PAC 10). The overall publication mean for all academic ranks combined was 5.9069 articles published over a five-year period.

In the PAC 10, instructors had the lowest overall mean publication rate of 2.3209 articles published in a five-year period as compared to all other academic ranks. This was also the case for instructors at O.S.U. The highest publication rate was noted for assistant professors in the PAC 10 and associate professors at O.S.U.

Full professors at Stanford U., U.S.C, and U.C.L.A. had the highest publication rate of all academic ranks. At these same institutions there appears to be a trend of increased publications as faculty move from one promotion level to the next, and most productive as full professors. The reverse is true at O.S.U., by the time faculty reach full professor status publication in referred journals ceases. The publication rates noted for W.S.U. and A.S.U. are lower than California institutions of higher education, but substantially higher at all academic ranks when compared to O.S.U.

Table 49

Faculty Publication-Rate Means by Academic Ranks for Colleges of Education Within Universities Belonging to the Pacific Athletic Conference (PAC 10)

Universities	Academic Rank Means					
	Inst.	Asst.	Assoc.	Prof.	overall	N
O.S.U.	0.0000	0.5714	1.2857	0.0000	0.5862	29
W.S.U.	1.6350	6.0667	5.5186	4.6356	4.6356	31
A.S.U.	1.9135	9.5319	6.2782	6.5244	6.0620	37
Stan. U.	3.1210	9.6344	9.9982	11.2253	8.3697	41
U.S.C.	2.7060	9.8901	10.3397	10.5860	8.1385	33
U.C.L.A.	2.2294	9.3076	9.1623	10.5860	7.8213	36
Means:	2.3209	8.8861	7.1137	7.3151	5.9069	

Table 50 depicts the inter-correlation among the 12 subscales of SATI. In the Huck, Cormier, and Rounds (1973) measure of central tendency and variability the magnitudes of a correlation coefficient values have been described as either being low or high negative or positive correlations. A third type was noted as the zero correlation in which no systematic relationship was said to occur between compared variables with magnitudes in the .00-.09 range. The closer the coefficient is to either +1.00 or -1.00, the higher or stronger the correlation is, and the closer the magnitudes of the correlations are approaching zero, the lower or weaker the correlation. These researchers provide the following examples of magnitudes of correlation coefficients in Pearson Product-Moment Correlations on page 31 of their book Reading

Statistics and Research:

+.95, +.85, +.93, +.87	high positive correlation
+.23, +.17, +.18, +.20	low positive correlation
+.02, +.01, .00, -.03	no systematic relationship
-.21, -.22, -.17, -.19	low negative correlation
-.92, -.89, -.90, -.93	high negative correlation

Table 50

Inter-Correlation Between Twelve Subscales of the SATI

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**
S1	1.0	.76	.77	.71	.72	**.34**	.56	.49	**.22**	.74	.76	.77
S2	.76	1.0	.82	.68	.71	**.31**	.51	.47	**.21**	.78	.77	.82
S3	.77	.82	1.0	.84	.79	**.31**	.60	.52	**.18**	.87	.81	.88
S4	.71	.68	.84	1.0	.78	**.37**	.73	.57	**.15**	.78	.76	.85
S5	.72	.71	.79	.78	1.0	**.34**	.63	.62	**.11**	.83	.84	.87
S6	**.34**	**.31**	**.31**	**.37**	**.37**	1.0	**.44**	**.44**	**.19**	**.33**	**.34**	**.35**
S7	.56	.51	.60	.73	.63	**.44**	1.0	.65	**.11**	.60	.62	.66
S8	.49	.47	.52	.57	.62	**.44**	.65	1.0	**.10**	.58	.65	.60
S9	**.22**	**.21**	**.18**	**.15**	**.11**	**.19**	**.11**	**.10**	1.0	**.14**	**.18**	**.16**
S10	.74	.78	.87	.78	.83	**.33**	.60	.58	**.14**	1.0	.82	.88
S11	.76	.77	.81	.76	.84	**.34**	.62	.65	**.18**	.82	1.0	.88
S12	.77	.82	.88	.85	.87	**.35**	.66	.60	**.16**	.88	.88	1.0

All correlations were positive and of a magnitude greater than .6 for over half of the 144 subscale inter-correlations, 55 of which had magnitudes greater than .7 of a point. Subscales #6 (The instructor provided scheduled office hours or was readily available for consultation with me) and #9 (The examinations were relevant to the reading assignments and to the material presented in class) revealed low to moderately low correlations to the other ten subscales.

CHAPTER V

DISCUSSION AND CONCLUSIONS

Introduction

The purpose of this study was to examine the relationship between select demographic variables and student evaluation scores of faculty teaching performance at a large Land Grant and Sea Grant University in Oregon. Most of the discussion that follows addresses the interaction between these variables.

Student evaluation results occupy a central role in the criteria used by administration to make faculty tenure and promotion decisions. As such, an investigation designed to generate empirical evidence on what effect demographic variables might have on student evaluations of faculty teaching performance would greatly facilitate summative and formative evaluation processes.

Chapter five is divided into four parts. First, the results of the data analysis are summarized. The discussion commences with the research data corroborating or contradicting the evidence revealed in the present study from hypothesis one to hypothesis thirteen of the study. Second, the implications of these results are discussed. Third, suggestions for future research are given. Finally, recommendations are provided to central administration on how best to use student evaluations for the purpose of improving and rewarding the teaching component of the tripartite role of faculty at Oregon State University.

Hypothesis 1

The null hypothesis could not be supported; significant differences were found between student ratings of faculty teaching performance in major and non-major courses. The purpose of this aspect of the investigation was three-fold. The first part involved determining if major and non-major courses influenced faculty overall rating results. The second part of the analysis required ascertaining if differences in student ratings exist between the college with the highest and lowest mean ratings on the major/non-major grouping along the 12 instructional dimensions of SATI. The final part of the analysis involved determining if faculty and student gender variables affect ratings in major course designates.

The results of the present study revealed that students in non-major courses tend to rate faculty higher than professors in major courses on the overall subscale of SATI. The analysis of variance resulted in an $F = 42.2278$ which was significant at the .05 level, and significantly larger than the theoretical table value of 3.84. The paucity of empirical research data available on this variable indicates a contrary view to the results of this study, e.g. no significant differences and no significant relationships were previously found between student ratings and major versus non-major courses (Aleamoni & Thomas, 1977; Null & Nicholson, 1972; Rayder 1968).

A plausible explanation, but certainly not the only one, for the significant difference in mean scores between these variables (major and non-major) might be related to student "motivational" factors. Centra and Creech (1976) argued that student motivation and student personal interest in major courses and chosen subjects of study would lead them to rate major courses as more valuable than non-major courses. Also, these researchers believe that teachers have less interest in non-major courses; therefore exert less effort in teaching these classes.

If motivational factors alone could account for the significance in the statistical comparison of these variables, then major courses should receive higher overall evaluation scores, such is not the case in the current study. Other student factors may be of importance in the interaction between major and non-major courses, i.e. attitudes generated from perceived levels of "risk" and "choice" students may have in major verses non-major courses. The same may be true in required verses elective courses as well. Students may be more critical of courses in their major compared to non-major for reasons not immediately reflected in global or overall ratings.

The results of this investigation suggest that students do, in fact, rate faculty differently in major and non-major courses overall. This was also evident at the college level of analysis. To explore where these differences and similarities occur along the effective teaching dimensions embodied in SATI, two colleges were selected for comparison based not only on extremes in mean ratings but also because of equivalence in the number of student responses.

A two-way ANOVA was conducted comparing the Colleges of Education and Pharmacy. The former consistently received high mean ratings on faculty teaching performance, while the later consistently received low ratings. This was evident on both the major/non-major and required/elective course variables on the overall subscale and at the college level of analysis.

The overall means from both colleges are statistically different in terms of magnitude of rating scores and levels of productivity. Productivity has been defined, for the purpose of this study, as publications rates in refereed journals. A difference in teaching performance mean scores greater than 0.4 was noted for non-majors, and greater than 0.7 for majors, and a difference of 12.4 on a scale ranging from 0 to 13 for productivity levels between the Colleges of Education and Pharmacy.

The number of responses in the major for each college were similar, e.g. Education, n= 667 (mean of 3.4276), and Pharmacy, n= 676 (mean of 2.7260), and were large enough for comparison. ROTC (mean of 3.5833) with n= 7 and Oceanography (mean of 2.7197) with n= 6, while they showed the largest and smallest overall faculty rating means (Table 7), they did not have a sufficient number of student responses to conduct an adequate comparison.

A difference greater than 0.7 was noted in the overall teaching performance mean scores for students taking courses in their major in the Colleges of Education and Pharmacy. Students in both colleges felt faculty were well prepared and organized when they came to lecture (subscale #2). Students also rated faculty in both colleges similarly along the instructor-student rapport dimensions of being sensitive to individual student/class ability to understand course material (subscale #4), and being available/willing for consultation after class (subscale #6). Faculty in both colleges stimulated enthusiasm for the subject matter taught in the course (subscale #5). They also were similar in their perceptions of having learned a significant number of new ideas/skills after exposure to the instructor (subscale #11). When asked to give an overall impression of faculty teaching performance, students in both colleges agreed that they were favorably impressed with the teaching performance of their instructors.

Faculty in both colleges were perceived by students as being well prepared and organized. However, significant perceptual differences greater than .6 occurred between students from the two colleges on (subscale #1) clear presentation of course objectives and requirements. This was also amplified in the clarity in which the course material was explained by faculty (subscale #3). Students in Education gave significantly higher ratings to their instructors on these rating dimensions than did students from Pharmacy. Apparently, it is possible for an instructor to be well prepared and organized in a content area but fail to make clear to students, course direction, expectations, and subject matter concepts.

Perceptions of student's best and worst large and small classes were collected on an open-ended questionnaire from 800 students at the University of Washington. Findings revealed that learning was assisted when instructors provided an overview of expectations at the beginning of each class, outlined major concepts, and made repeated emphasis on main ideas. This was especially important to students in large classes (Wolff, Nyquist, & Abbott, 1987).

In this study students also rated their satisfaction in their best large class significantly higher than their best small class when instructors gave examples and illustrations, more specifically, "using a large number of examples pertaining to the material" and "using examples everyone can relate to" (p. 25). Particularly useful to students learning was the use of demonstrations and analogies that reinforce concepts. When this was done with humor the content was perceived as being more engaging. "Boring professors" and "boring lectures" were noted as major factors that hindered the learning process (Wolff et al., 1987).

In the present study the largest difference in overall mean scores between non-majors in Education and Pharmacy was found on the instructional dimension of relevance of examinations to reading assignments and subject matter presented in class (subscale #9), a difference greater than 0.7 of a point on a five-point scale. Students in Pharmacy perceived the objectives of assignments as having little connection with the content covered on examinations.

In the Colleges of Pharmacy and Education student perceptions on subscale #1 and #9 showed a low positive correlation of $r= .22$, but a high positive correlation of $r= .77$ between subscales #1 and #3. Perhaps a clear understanding of expectations in course objectives and requirements also influenced student perceptions on how clearly faculty explained material in class (subscale #3), but not what student's thought instructors would emphasize from class assignments on examinations (subscale #9).

Clear expectations may have forced students to forgo independent thought for a focus on determining what faculty would cover on examinations. Pharmacy students noted on subscale #8 that they did not perceive instructors as encouraging them to think for themselves, a positive correlation of r= .49 was found between student perceptions on subscales #8 and #1.

In the College of Education mean rating scores on all subscales of SATI for non-majors were, on an average, 0.4 higher than the mean teaching performance rating scores noted for the College of Pharmacy. The ratings given to faculty in Pharmacy by students at the undergraduate level of analysis (hypothesis #6) also revealed major differences in instructional effectiveness in the graduate/undergraduate level of analysis. Education faculty are more readily available for student consultation than Pharmacy faculty as perceived by undergraduate students.

Two-way ANOVA was conducted on 27,400 student responses comparing faculty ratings based on student gender and major/non-major courses (Table 8). Results reveal that no interaction was evident in the overall and main effects for 11 of the 12 subscales on SATI. Subscale #7 (The instructor was fair and impartial in dealing with me) was significant at 0.05 level of confidence for female students rating in major (3.7) and non-major (3.2) courses. A mean difference of 0.5 of a point on a five-point scale was noted on this subscale rated by female students. The correlation coefficient on subscale #7 and nine other subscales was greater than mean correlations coefficient .5 for them all.

Female students might be more sensitive to this instructional dimension than male students, as was discovered on other instructional dimensions in other studies. Elmore and LaPointe (1974, 1975) found that female students appeared to be more sensitive to the structure aspect of effective teaching than were male students. McKeachie, Lin, and Mann (1971) have shown that instructors who were rated high on course structure were also more effective with female students.

Hypothesis 2

The null hypothesis was unsupported as significant differences were found between student ratings of faculty teaching performance in required and elective courses. The purpose of this aspect of the investigation was two-fold: first, to determine what influence required/elective variables had on the overall subscale of student rating of faculty teaching performance; and second, to determine if relationships exist in the general variable association between the required/elective variable grouping and the major/non-major variable grouping on any 12 SATI instructional dimensions.

The results of this investigation suggest that students do rate faculty differently in required and elective courses on the overall subscale. Several investigators corroborate the findings of this study for the variable of required verses elective courses on overall rating subscales. Gillmore and Brandenburg (1974) and Pohlmann (1975) found that students who are required to take courses tend to rate them lower than students who elected to take courses. These findings are supported by Gage (1961) and Lovell and Haner (1955), who found that professors of elective courses were rated significantly higher than professors of required courses. They argue that instructors may operate more effectively in elective courses and that students in these classes are easier to please because they are better motivated; no other explanations or conjecture was posited.

In contrast Hildebrand, Wilson, and Dienst (1971) reported no difference between the ratings of students in required and elective courses. In the Gillmore and Brandenburg (1974) study, the unit of analysis was the course section with no reference to colleges. The authors noted an overall mean score of 3.26 for elective course sections having 60% to 79% of their students taking courses as a requirement (21% to 40% elective). This mean evaluation score closely parallels the overall mean score of 3.24 in the current study for elective courses represented by 31.17 percent of the students in the elective group.

Pohlmann (1975) also noted that the percentage of students in courses taken as an elective correlated positively ($r = .27$) with the general course rating variable. Additionally, as the percentage of students electing to take a particular course increased, so did the overall evaluation scores for the course itself.

Lovell and Haner (1955) found a statistical difference between required and elective course variables on the global subscale. Instructors teaching required courses received significantly lower rating than those teaching non-required courses. These researchers while examining the effects of class size on student rating results, noted that the highest mean ratings were received by instructors with classes of 21 to 30 students, concluding that these courses were not required for graduation.

Classes with student numbers ranging from 1 to 10 received the second highest mean score ratings. Classes with 31 or more students received the lowest mean rating scores, most of these were required. While not expressly stated, the required and elective variable appeared to confound the class size variable resulting in a statistically significant difference between large and small class sizes, where in fact a statistical difference may not have existed among the class size variables. In the present study no significant difference was found between small or large classes based on the Minimum Class Size Policy at Oregon State University.

The pervasive nature of the elective course variable on faculty teaching performance scores can be seen at the college level of analysis. In table 10 all colleges (with one exception, the College of Liberal Arts a mean difference of 0.15) showed mean scores higher for students electing to take courses as opposed to those students taking courses as a requirement.

Between colleges this difference is also quite noticeable. In the required course variable the overall rating scores ranged from a high in ROTC of 3.5329 to a low in Pharmacy of 2.6475 with a mean difference of 0.8854 of a point. In the elective course variable the overall rating scores ranged from a high in ROTC of 3.6890 to a low in Science of 3.0459 with a difference of means of 0.6431 on a five point-scale.

In the present study faculty teaching in elective and non-major courses received the highest overall ratings. Although, among colleges the range in mean scores for the elective and non-majors variables was less than 0.5 of a point, while the range in overall mean scores for the required and major variables was greater thab 0.8 on a five-point scale.

Two-way ANOVA results comparing major/non-major variables with required/elective variables showed no interaction on any of the 12 subscales of SATI. However, main effects revealed the required/elective course variable grouping to be the most influential on ratings faculty received from students. Within this grouping mean ratings were significantly higher for faculty teaching elective courses, meaning students perceived these faculty to more effective teachers, especially on seven of the twelve SATI subscales.

Students perceived faculty in elective courses to be better prepared and organized, clearer in presenting course material, in stimulating more enthusiasm for the subject matter of the course, and being more effective communicators. These same faculty were also perceived by students, in the low inference domain of instructor/student rapport, to be the most effective, e.g. more concerned that students and the class understood concepts, in dealing fairly with students, and in showing more impartiality.

The greatest difference among the four variables combined was on the relevancy of examinations to the reading assignments and material presented in class. In the present study, this instructional dimension (examination subscale #9) would continue to show that it had a negative influence on student perceptions of faculty teaching performance ratings regardless of what demographic variables were compared, faculty rank or the gender of instructor or student.

Main effects in the five remaining subscale (#1, #6, #8, #11, and #12) were due to variable(s) other than the required/elective and major/non-major variable groupings. Although, the mean ratings for elective/non-major grouping were significantly higher than the required/major variable grouping on all five subscales.

The elective/non-major the elective course variable was the most influential on these same five subscales. In other words, on the elective variable, students perceived faculty course objectives and requirements to be more clearly presented, and faculty were perceived as being readily available for consultation and encouraged students to think independently. Whatever variable(s) were responsible for the main effects on these five subscales resulted in students perceiving faculty teaching elective courses to have availed them with a significant number of new ideas/skills, and a favorable impression of faculty teaching competence. Faculty teaching non-major/elective courses received mean ratings significantly higher than faculty teaching elective and non-major courses on all SATI subscales.

Results comparing required/elective variables on the instructional dimensions of SATI indicate significant differences exist on student perceptions of effective teaching by whether in required and elective courses at p .05 level of confidence on four twelve subscales. These also overlapped interactions found on the major/non-major variable grouping.

Hypothesis 3

The null hypothesis could not be supported, as significant differences were found between faculty teaching performance means and student grade point average (GPA). The purpose of this aspect of the study was to determine if a relationship was present for levels of the GPA student variable and the means of faculty teaching performance ratings. Additionally, the GPA variable was compared to the expected grade level variable to determine if a relationship existed.

No attempt was made in the present study to investigate the accuracy of the actual and reported GPAs of students on the Student Assessment of Teaching Inventory (SATI). In an extensive review of over 200 articles Feldman (1976) noted that the actual and reported GPAs were highly correlated. Not only did students remember their GPAs but they were also honest in the reporting them as noted on course evaluation forms. Rayder (1968) investigated the validity of student information in the reporting of GPAs and found a correlation of .96 between actual and reported GPAs Students were able to remember and report their GPA accurately.

In the present study no interaction was found in a two-way ANOVA at a confidence of .05 comparing the independent variables of GPA and expected grade. A positive relationship was revealed between the overall mean scores of student GPA and the overall mean scores of expected grade. Students who reported the highest GPAs and expected grades also gave the highest faculty teaching performance ratings. Conversely, student reporting the lowest GPA gave the lowest faculty ratings and expected to receive the lowest course grade.

Hypothesis 4

The null hypothesis could not be supported, as significant differences were found between faculty teaching performance means and student expected grade. Considerable controversy has centered around the relationship between student ratings and their expected course grade. Some researchers have gone so far as to say that students tend to rate courses and instructors more highly when they expect to receive good grades.

No attempt was made in the present study to acquire actual grades received in the course for correlation with expected grades for 40,000 student responses. However, Feldman (1976) noted that the actual and expected grades were highly correlated. What students reported as an expected grade closely matched the actual course grade received.

In the present study findings showed that student evaluation scores of faculty teaching performance do, in fact, closely parallel the grades students expected to receive at the end of the course. Students expecting to receive high grades gave the highest overall faculty evaluations. Students expecting the lowest course grades gave the lowest overall faculty ratings. The H or Audit grade level had the highest overall mean faculty teaching ratings. Previous research provides no evidence for the H grade designate; hence why students auditing courses should give the highest ratings on faculty teaching performance.

In the present study it was not known if students were aware of their final course grades prior to the time they filled out the faculty evaluation of teaching form. It is the belief of this researcher that the majority did not. In the investigation by Brown (1976) course grades were known to students when they filled out the course evaluation forms and were correlated with faculty ratings. Foreknowledge of course grades made a significant difference in faculty teaching performance rating outcomes.

The results of the present study suggest that ratings of faculty teaching performance are influenced by the grades students expect to receive in their courses. Not only was the expected grade variable positively related to faculty teaching performance ratings at all levels, but two-way ANOVA results revealed that they were also positively related to all levels of the reported GPA student variable as well. This is contrary to the results of all other investigations.

Feldman (1976) in a review of over 200 studies concluded that "currently available evidence cannot be taken as definitely establishing a bias in teacher evaluation due to grades students receive or expect to receive in their courses, but neither is it presently possible to rule out such a bias" (p. 69).

Scheurich, Graham, and Drolette (1983) analyzed expected grade as one of ten instructional dimensions of effective teaching and concluded that expected grade was the least important predictor of teacher ratings. Even though expected grade was the least important predictor, it showed a correlation value of .48 with the other nine rating dimensions, eight of which are found on SATI. These findings and the magnitudes of the correlations were consistent with the studies summarized by Feldman (1976), Stumpf and Freedman (1979), Stumpf (1979), and Freedman and Aguanno (1979).

The H or audit grade level showed the highest overall ratings. Students auditing courses were not responsible for class assignments or course examinations nor did they have the pressures and assignment deadlines associated with graded classes that would perhaps force them to be more aware/critical of the quality of instruction that was provided. Without these attending factors, is it possible for auditing students to be more objective and/or comprehensive in their faculty ratings than for students taking these same courses for grades? This researcher thinks not, especially when one considers that the important effective teaching dimension of evaluation--determining student understanding of subject matter concepts, the rationale on which conclusions were based, and the feedback necessary to correct misperceptions--was missing in the audit courses.

As one might expect the majority of the audit students had "no basis for opinion" on the SATI subscales #9 (the examinations were relevant to the reading assignments and to the material presented in class). The few that responded to this subscale had no basis for their response unless they had taken the course examinations; there was no easy way of making this determination.

Hypothesis 5

The null hypothesis could not be supported, as significant differences were found between student percent of the class attendance variable and ratings of faculty teaching performance. The purpose of this aspect of the investigation was two-fold: first, to determine if a statistical relationship was present between faculty ratings and how often students attended a particular class in which they rated during the course; and second, to ascertain if percent of the class attended was significantly related to the student variables of reported GPA and expected grade.

One-way ANOVA result revealed significant differences between faculty ratings and the percent of class attendance by students. Students who attended class more often during the course gave higher faculty evaluation scores. These students had more contact hours with faculty members; therefore it was possible for them to form a more accurate and clearer perception of the classroom behavior manifested by faculty over a longer period of time in which to make an assessment of effective teaching.

Aleamoni (1976) uses this argument for the inclusion of student evaluations in summative evaluation processes as being just as valid an indicator of actual faculty teaching performance as is faculty peer ratings. His argument was based on the longer observation of instruction contact hours for students.

Three-way ANOVA results showed no interaction between percent of class attended, reported GPA, and expected grade. A difference was found at the .05 level of confidence for the percent of class attended variable. Positive relationships exist between faculty teaching ratings and percent of class attended, reported GPA, and expected grade at all levels of these variables. Students who attend class more frequently gave higher faculty ratings, reported higher GPAs, and expected higher grades.

Hypothesis 6

The null hypothesis could not be supported, as significant differences were found between student ratings of faculty teaching performance across student class level and graduate/undergraduate courses. Univariate analysis indicated that graduate student evaluation scores of faculty teaching performance were significantly higher than those discovered for undergraduate students ($F= 92.0313$, $p= .0000$).

The significant difference between graduate and undergraduate courses was shown to occur at the freshman level. The mean ratings for 200-400 level students were found to be statistically indistinguishable from each other. The means of the freshman (100 level) and graduate (500 level) students were statistically different from each other and from the ratings of 200-400 level students.

Opposite results were obtained by Aleamoni and Graham (1974) within the course level variable using multivariate analysis. The purpose of their study was to determine if members of higher rank received the highest student ratings when class size and course level were taken into account in a sample of 448 course sections. Multivariate analysis (MANOVA) results of class size, course level, and instructor rank indicated highly significant differences on rating scores at the course level, but no interaction was discovered for the class size variable compared singly by the six subscales between small (1-20), medium (21-40), and large (over 40) course sections or by academic rank from instructor to full professor.

In the above study instructors in freshman and graduate courses were rated lowest on the subscale general course attitude, while instructors in junior- and senior-level courses received the highest ratings. This sub-scale matches the global or overall subscale of the Student Assessment of Teaching Instrument (SATI) used in the present study, which revealed freshman and graduate students gave the highest ratings while and sophomores, juniors and seniors gave the lowest ratings to instructors teaching those classes.

Aleamoni and Graham (1974) argued that the multivariate approach was preferable to conducting a series of univariate analyses of bivariate data in order to determine relationships in certain areas of interest. For this reason 3 X 5 X 5 MANOVA for 12 dependent variables were conducted on the data set in the present study. The results showed interaction at the class level on all 12 SATI subscales. Freshman and senior students gave the highest mean ratings on faculty teaching performance and sophomore and junior levels gave the lowest ratings.

In examining the differences in rating means between and within colleges for the undergraduate/graduate variable, only one significant result was found. The College of Pharmacy showed a difference greater than 0.8 between graduate and undergraduate overall mean ratings. For other colleges the within difference was less than 0.3. The range of rating scores between colleges for graduate students was from a high of 3.6074 for the College of Pharmacy and a low of 3.0107 for the College of Business, a difference of 0.5967. The undergraduate range from a high of 3.6201 for ROTC to 2.7235 for the College of Pharmacy, a difference of 0.8966 on a five-point scale.

The difference between undergraduate and graduate ratings of faculty teaching performance for the College of Pharmacy was shown to be significantly below university standards for undergraduate instruction with an overall mean ratings of 2.7235. However, graduate students rated Pharmacy faculty higher on effective teaching than did graduate students from all other colleges compared in the study.

The Colleges of Science and Engineering were also found to be low on undergraduate instruction. Low ratings at the undergraduate level may be a function of perceived academic rigor. Two-way ANOVA statistic was used to explore where these difference occurred along the 12 instructional dimensions (subscales) of SATI.

Interaction was discovered on eight dimensions, at the .05 level of confidence. The differences in mean rating scores between the graduate and undergraduate variables was greater than .7 on a five-point scale for all nine subscales. In his book, College: the Undergraduate Experience in America, Ernest Boyer (1987) argues that competent and dedicated teaching is the central focus of the undergraduate experience and that all faculty "should work continuously to improve the content of their courses and their methods of instruction" (p. 159).

The first of the eight interactions was found on subscale #1 (Course objectives/requirements were clearly presented). Undergraduate students seemed not to know what to expect from their courses as stated by faculty who taught at this level. This perception may have emanated from a general sense that faculty were not well prepared and organized (subscale #2), having received a mean rating of 2.5631 on this instructional dimension; a correlation value of .76 was found between these two variables. If course expectations are ambiguous and lectures characterized by disorganization (r= .77), then one could expect the clarity in which course concepts were explained in class (subscale #3) to be as deficient, but it was not, even though the faculty received a low mean rating of 3.1146 in the main effect on this dimension due to undergraduates. Graduate students also rated faculty low but above university standards.

Faculty were also perceived by undergraduate students to lack sensitivity to individual student and the classes' ability to understand course material (subscale #4) which may in part explain the low rating received on subscale #3; a correlation of .84 was found when these subscales were compared. The lowest ratings given by students was on instructional dimensions #5 (The instructor stimulated enthusiasm for the subject matter of the course), #8 (The instructor encouraged me to think for myself), and #9 (The examinations were relevant to the reading assignments and to the material presented in class). The ratings faculty received on these three instructional dimensions were far below university standards of instructional effectiveness.

Ernest Boyer (1987) captures the essence of the quality of undergraduate instruction as it pertains to the three instructional dimensions just mentioned in the following comment:

> The central qualities that make for successful teaching can be simply stated: command of the material to be taught, a contagious enthusiasm for the play of ideas, optimism about human potential, the involvement with one's students, and--not least--sensitivity, integrity, and warmth as a human being. When this combination is present in the classroom, the impact of a teacher can be powerful and enduring. Good teaching also means careful evaluation of the student. (p. 154)

When these qualities are lacking students may perceive their learning experience as not availing them with a significant number of new ideas and skills (subscale #11), and may not be favorably impressed overall by their instructors (subscale #12), as was evident in the perceptions of the undergraduate students. Main effects were found on subscales #6 (The instructor provided scheduled office hours or was readily available for consultation with me), subscales #7 (The instructor was fair and impartial in dealing with me), subscales #10 (The instructor used good communication skills), all due to the undergraduate level with faculty rating means significantly higher as given by graduate students.

No differences occurred in undergraduate students' perception of faculty teaching effectiveness in Pharmacy with faculty in the Colleges of Science and Engineering on subscales #1, #4, #9, #11, and #12. Again, students noted faculty instructional deficiencies in the areas of knowing course expectations, feeling like faculty cared whether they understood the course material, and germane examinations. As with the College of Pharmacy, students in the Colleges of Science and Engineering noted not having learned a significant number of new ideas/skills at the end of the course and were not favorably impressed with the faculty who taught these courses.

Interestingly, these three colleges were found to have the highest productivity rates (publications in refereed journals) compared to all other colleges in the study and the lowest undergraduate ratings of effective teaching in the graduate/undergraduate variable analysis. Although, a Pearson product moment correlation result hypothesis #13 showed a correlation of -.0293 for faculty ranks overall. Low overall coefficients indicate that no signicant relationship exists between faculty publication rates and mean ratings of teaching performance for all academic ranks.

Hypothesis 7 and 8

The null hypothesis could not be supported, as significant differences were found between male and female students' ratings of male and female faculty teaching performance. Two-way ANOVA results revealed significant differences in student rating scores of faculty teaching performance; therefore the null hypothesis was rejected. The value of $F= 92.0313$ was significantly larger than the table value of 3.84, at a $p= .05$.

This supports the findings of other investigations where student gender interacted with faculty gender to effect faculty ratings (Kaschak, 1978; Lombardo & Tocci, 1979; Basow & Silberg, 1987). Depending on the college and gender of student/faculty within the college significant differences in student rating scores of faculty teaching performance were found on a number of the instructional dimensions of SATI. These significant differences in faculty ratings appeared in two colleges (Home Economics and Engineering), and they were compared for analysis on SATI subscales.

The purpose of this part of the study was to determine if gender of the student interacted with gender of the instructor to influence teaching performance ratings across 12 instructional dimensions of SATI. Hypotheses 7 and 8 were combined to facilitate an understanding of this interaction. Consequently, for specific data analysis and interpretation they could not be separated as dependent and independent variables and produce a meaningful data result capable of being compared to other research studies.

Certain disciplines are thought of as being characteristically male dominated, i.e. Engineering; whereas others are thought of as female dominated as in the College of Home Economics. In the present study male and female faculty were not directly matched on the basis of faculty age, rank, years of teaching experience, seniority or tenure. Feldman (1983, 1986, 1988) in an extensive review of the research results of over 300 articles pertaining to student evaluation of faculty teaching performance, concluded that these variables had little or no influence on faculty evaluations.

Investigations have shown that faculty teaching performance ratings within certain disciplines and on certain instructional dimensions have been influenced by gender stereotypic perceptions and bias of students toward faculty who did not fit accepted gender patterns or career fields (Etaugh & Riley, 1983). These gender patterns were described as masculine (instrumental/agenetic), feminine (expressive/warm), or androgynous. Female faculty sometimes received lower teaching performance ratings, especially from male students, than were received by their male counterparts. Gender-stereotypic characteristics influenced evaluations of college faculty (Basow & Distenfeld, 1985; Basow & Howe, 1987; Basow & Silberg, 1987; Erdle, Murray, & Rushton, 1985; Holahan & Stephen, 1981; Bennett, 1982; Harris, 1975,1976; Kaschak, 1978, 1981).

In the present study results indicate that female faculty have significantly higher overall teaching performance ratings when compared to the overall ratings discovered for male faculty. This is contrary to the findings of Kaschak, (1978) and Lombardo and Tocci, (1979) where female professors consistently received less favorable overall ratings by male students on all measures of effective teaching except on the instructor/individual student interaction dimension. This instructional dimension is related to a professor's availability for student contact in class and outside class for consultation purposes.

Bennett (1982) argued because of gender stereotypes, female faculty were possibly expected to be more accessible to students than expected of male faculty. Basow and Silberg (1987) found that both male and female students rated female professors less favorably on the instructor/individual student interaction and conjectured that these results may be attributed to their not having conformed to these expectations.

Main effects were found in the present study on the instructor/individual student interaction dimension of SATI denoted by subscales #4 (The instructor was sensitive to my/the class' ability to understand the material) and #6 (The instructor provided scheduled office hours or was readily available for consultation with me) due to faculty gender which revealed female faculty had received significantly higher ratings on these instructional dimensions than did their male counterparts. Whether these results are due to female faculty meeting the expectations described by Bennett (1982) or because they are just more effective at this level of instruction than male faculty is unknown. Further research is needed concerning student ratings of male and female faculty on this instructional dimension coupled with the actual time spent by each group fulfilling this student need.

Main effect on subscale #4 of SATI (The instructor stimulated enthusiasm for the subject matter of the course) was also due to instructor gender, not student gender. In this case female faculty, as in the results of subscales #4 and #6, received significantly higher ratings than did male faculty. Thus female faculty were perceived as being more effective, not only in their personal contact with students in and out of the classroom, but in stimulating student interest in the subject matter as well.

Basow and Silberg (1987) reported that female faculty were rated lowest on the dynamism/enthusiasm dimension by male and female students. They attributed this finding to the strong correlation ($r = .88$, $p < .001$) between the overall teaching ability, favoring male professors, and the Dynamism/Enthusiasm dependent variable. In the present study the correlation between the overall ratings of effective teaching on SATI, favoring female faculty, and the enthusiasm variable (subscale #5) was .76, $p > .05$. In either study direct observations of faculty classroom behavior was not conducted. If faculty were engaged in different classroom behaviors it would be important to know what they are and how they are related to teaching performance ratings as a function of faculty and student gender.

In Figures 1 and 2 interaction was found and plotted on the remaining nine dependent variables on SATI between gender of the student and gender of the instructor (#1, #2, 3#, #7, #8, #9, #10, #11, and #12) for the Colleges of Engineering and Home Economics. These were the same nine instructional dimensions found on the overall analysis of student/instructor gender compared on the 12 subscales of SATI. In studies where the influence of gender of student and faculty is analyzed in terms of faculty ratings results, the number of male faculty represented in home economics and female faculty in engineering tends to be low. The present study is no exception; however, meaningful data can still be derived if caution is exercised in the interpretation of these data with the small sample size in mind.

In general, female students' ratings of female faculty revealed the smallest sum of the mean difference among the four student/instructor groupings. A difference of a mean of 0.1268 on a five-point scale was noted between these two colleges. Female students rated female faculty similarly from one subscale to the next on SATI and independent by one college to the next across the university.

This corroborates, in part, the results of Kaschak (1978, 1981) where overall female student ratings of female faculty were consistent across disciplines. This consistency in evaluations included male faculty as well. In the present study, however, the widest range in teaching performance scores across most instructional dimension was due to female students, not male students, ratings of male faculty; and the results were college dependent. In contrast, male students in the Kaschak studies showed a consistent bias in favor of male faculty. Both student genders were noted as more willing to take courses from faculty of the same gender.

In the present study this gender preference was also evident in the analysis of male student ratings of male faculty, although in non-traditional disciplines. Male students rated male faculty significantly higher than female faculty in Home Economics on all subscales. However, in the College of Engineering male students rated female faculty significantly higher on all subscales than female students ratings of female faculty in the same college (Figures 1 and 2).

In the Kaschak (1978) study male students consistently rated the teaching methods of male professors as being more effective, concerned, likeable and excellent. The later study by Kaschak (1981) male professors were still characterized in their teaching as being more powerful and effective than female professors. Female professors in "feminine" fields were rated as more concerned and likeable than their male counterparts.

In some studies male faculty in home economics and female faculty in engineering were perceived by students as not fitting an acceptable gender related field as was reflected in their teaching performance ratings, and descriptions of appropriate career fields (Etaugh & Riley, 1983). Basow and Distenfeld (1985), and Basow and Silberg (1987) found that male and female students majoring in engineering gave less favorable ratings to female professors than to male faculty. However, the results in the present study suggest just the opposite conclusions in that student ratings favored faculty in nontraditional disciplines.

On all SATI subscales the mean ratings given by male students to female faculty were significantly higher in Engineering than for male faculty; whereas in Home Economics female students rated male faculty higher than female faculty on six subscales and equivalent to female faculty on three subscales. In fact, the highest mean rating in the data set were given by female students to male faculty in Home Economics, as well as the lowest to male faculty in engineering. These mean differences were significant at the .05 level and in some cases at the .001 level.

Male students' rated male faculty significantly higher in home economics than did female students ratings of female faculty in the same college on all subscales including the overall (subscale #12). The same was found on all instructional dimensions for female student ratings of female faculty in the College of Engineering which were higher on all subscales than male students ratings of male faculty.

The differences between the means for the Colleges of Engineering and Home Economics on the nine SATI subscales was 0.4 in the female student/male faculty grouping favoring male faculty in home economics. This difference represented the largest difference in means found among the four student/faculty groupings, although small by comparison to differences in the means found in female student's ratings of male faculty on most of the instructional dimensions embodied in SATI subscales.

Female students perceived male faculty in Engineering as having below average ability to clearly communicate course objectives and requirements (subscale #1). They gave male faculty a mean rating score of 2.9959 on this subscale, which is below university standards. A mean rating difference of 0.0041 on a five-point scale was found between male faculty in Engineering and Home Economics on this subscale.

The low rating on subscale #1 may, in part, explain why female students perceived male faculty in Engineering as having a below average ability in communicating course material clearly (subscale #3), but this would not explain why male faculty in Home Economics, who received a mean rating score of 3.6667 by female students on subscale #3, should also receive a low rating on subscale #1. If male faculty in Home Economics were perceived by female students as being effective in conveying course concepts, then continuity of perceptions should hold that the communication style employed to convey course concepts, would also be perceived by female students as equally effective in presenting course objectives/requirements clearly throughout the course, but it was not.

Elmore and LaPointe (1974, 1975) found in their studies that the only significant difference between the ratings of male and female faculty was on the rating item where the instructor specified objectives of the course. Female students rated female instructors significantly higher on course organization/expectation. Female students appear to be more sensitive to the structure aspect of effective teaching than are male students. McKeachie, Lin, and Mann (1971) have shown that instructors who were rated high on course structure were also more effective with female students.

Male faculty in Home Economics were perceived by female students as being more effective communicators of course expectations and course concepts and fairer/more impartial in handling students than male faculty in Engineering. As a consequence, one might expect to see an equally high mean ratings for male faculty in Home Economics on the relevancy of examinations (subscale #9), and on learning a significant number of new ideas from the course (subscale #11). However, male faculty in both the Colleges of Home Economics and Engineering received only average ratings on both subscales by female students.

Male faculty in both colleges were given a mean rating of less than or equal to 3.0 on subscale #11 (As a result of having this instructor, I have learned a significant number of new ideas and/or skills). The ambiguity in perception of course objectives may have contributed to the misperceptions held by male and female students that male faculty in Engineering did not explain the course material clearly. It may have also influenced female student perceptions on subscale #9 (The examinations were relevant to the reading assignments and to the material presented in class) and why male faculty received only an average mean rating of 3.0 on this subscale.

Male students also rated male faculty low (2.8696) on the clear presentation of course material (subscale #3). However, male students rated male faculty significantly higher than female faculty in Home

Economics on this subscale. The preference for male faculty by male students was dependent on the college being rated; the same results were found for female students.

Further evidence in the present study showing that male and female student ratings favored faculty in non-traditional disciplines comes from an analysis of how students perceived they were treated by faculty, what they perceived to have learned from the course, and their overall impression of faculty instructional effectiveness. Female students gave male faculty in Home Economics the highest rating of 4.0 in the data set on subscale #7 (The instructor was fair and impartial in dealing with me), whereas female students gave female faculty a mean rating of only 3.4609 in the same college.

Why male faculty in Home Economics were perceived by both male and female students as being more fair and impartial than their counterparts in Engineering is uncertain. Although in the major/non-major variable analysis only this variable was significant at the .05 level of confidence for female students taking courses in their major area of concentration.

Two-way ANOVA was conducted on 27,400 student responses comparing faculty ratings based on student gender and major/non-major courses revealed no interaction on 11 of the 12 subscales of SATI. A main effects was discovered on the sex of the student variable for the dependent variable (subscale #7) at the .05 level of confidence. Female students gave female faculty a mean rating of 3.7 and for non-majors a rating of 3.2 points on a five point scale. A mean difference of 0.5 of a point was noted on this subscale rated by female students.

Sensitivity to equitable treatment in a course may be just as important to female students as was course structure noted by Elmore and LaPointe (1974, 1975) and McKeachie, Lin, and Mann (1971). Kaschak (1981) argued that male faculty in a non-traditional discipline, e.g. Home Economics, may be perceived as a "novelty"; hence, a transposition of role preference that would ordinarily be the domain of female faculty.

Male students gave female faculty 3.5560 and male faculty 3.2265 mean ratings in Engineering on subscale #7. Male students also rated male faculty below university standards on three of the nine subscales which may have contributed their perception that they did not learn a significant number of new ideas/skills from having a male instructor (subscale #11) and were not favorably impressed by having had male instructors (subscale #12).

Male faculty in Home Economics, in contrast to Engineering, were perceived by female students as being more effective communicators in explaining course material (subscale #3), well prepared and organized in class (subscale #2), encourage students to think for themselves (subscale #8), and possessed good communications skill (subscale #10), having received a 3.6 mean rating on all four subscales. As a consequence, one might expect to see equally high mean ratings for male faculty in Home Economics on the relevancy of examinations (subscale #9), and on learning a significant number of new ideas from the course (subscale #11). However, male faculty received only an average ratings of 3.0 on both subscale.

In summary, the present study supports the importance of gender of the student and gender of the instructor in the evaluation of college faculty. On the three subscales with main effects, denoted as the instructor interaction variable (subscales #4 and #6) female faculty received from both male and female students the highest overall ratings. This was also true for the enthusiasm subscale (#5). If female faculty are expected to outperform male faculty on the first two subscales already mentioned by virtue of their gender as suggested in the literature, then they are doing exceedingly well. Just how much time and effort were expended by female faculty versus male faculty in this area is unknown but should be investigated further in subsequent research to throw light on the effects of these variables on faculty ratings.

Female faculty did receive significantly higher overall teaching performance ratings for all colleges combined as compared to their male counterparts. The magnitude of these ratings, however, were dependent on student/instructor gender in colleges in which courses were taken/taught.

Students gave significantly higher ratings to faculty of the opposite gender who were teaching, in what some researchers have called, non-traditional disciples. Whether these ratings are due to a gender preference or the demonstration of classroom behaviors as a function of teaching ability and/or gender was not investigated.

If student and faculty perceptions of classroom teaching were known, compared, and presented to faculty for the improvement of instruction, then maybe some of those behaviors might be identified and subsequently modified in the context of actual classroom teaching. Some of these behaviors may be found to be gender dependent; hence out of the control of individual faculty, in which case adjustments in rating would have to be made. These adjustments may have to be done statistically on teaching performance rating scores or in the degree of emphasis in summative evaluation processes.

Hypothesis 9

The null hypothesis was supported as significant differences were not found between student ratings of faculty teaching performance across the class size variable. The purpose for this analysis was to investigate the general assumption that as class size increases there is a corresponding decrease in quality of instruction and student course satisfaction. A claim frequently made by faculty and some administrators is that instructors of large classes receive lower ratings because students prefer small classes which permits more student-instructor interaction.

Univariate analysis was conducted in four ways: (1) based on the Minimum Class Size Policy established by Oregon State University, (2) natural divisions in the data set, (3) according to a number above and below 12 students per class, and (4) based on the comparison of class size with an instructor's teaching effectiveness determined by certain instructional dimensions recommend in the study of Wulff, Nuquist, and Abbott (1987).

Aleamoni and Graham (1974) suggested that further research should be conducted on data grouping in the class size variable to determine what factors influence student ratings of faculty teaching performance. Wulff et al. (1987) extended this suggestion to include analysis of class size categories in terms of student perceptions of class size and their subsequent influence on certain organizational/clarity instructional dimensions. The results obtained in various investigations may have been influenced by the classification schemes used to divide the data into logical units for analysis.

The minimum class size policy was based on a quality of instruction rationale and on economic considerations. The appropriate number of students at the appropriate level would serve both the interests of the students for quality instruction and the institution by sound management of fiscal resources.

To reiterate from Chapter 3, the Minimum Class Size Policy set by the Registrars Office at Oregon State University recommends that the minimum number of students allowed per undergraduate and graduate classes should be 16 students in lower division courses, 12 in upper division courses, and eight for the graduate level. In the current investigation student numbers below these benchmarks represent small classes and above these benchmarks represent large classes. The total number of classes was 786 for the small class size and 947 for the large class size with a total of 1733 classes in the study.

Results from the minimum class size grouping were not significant suggesting that student assessment of faculty teaching performance ratings are not influenced by class size. Lower- and upper-division courses and graduate courses showed no significant difference in the mean ratings given to faculty. However, two-way ANOVA analysis results indicated interaction on the organizational/clarity dimensions of subscale #1, #2, #3, #4, #5, and #11, with an overlap of subscales #9, #10, at the .05 level of confidence. The overall ratings masked differences in student perceptions of effective teaching noted by instructional dimensions on individual SATI subscales.

Students rated their large classes significantly higher than their small classes on the dimension of course objectives and requirements clearly presented (subscale #1). Having an overview of expectations at the beginning of and during the course was valued most by students in large classes. This might include an emphasis on important major concepts with enough repetition (review and summary) to maintain continuity from one lecture session to the next and/or to integrate laboratory work with lectures.

Also important to students in large classes were the degree of organization and clarity of instructors (subscales #2, and #3). Students preferred instructors who came to class well prepared, e.g. had organized lectures (subscale #2) and presented the content thereof with a high degree of clarity (subscale #3) and with good communications skills (subscale #10) so that most students could understand and would have little problem following the development of main concepts. According to Wolff et al. (1987) a characteristic of a well prepared and organized instructor in a large class was one that consistently followed a familiar instructional format paralleling the format presented in the text and syllabus.

Students in large classes were also sensitive to the relevancy of material presented in class and reading assignments as reflected in examinations (subscale #9) that emphasized what the instructor thought was important test material, which was not entirely perceived the same way by students in large class sizes. Student sensitivity to the examination in the large class size variable might in part be explained by student perception of the instructor-student interaction dimension.

The greatest difference between rating means of large and small classes on SATI, and the only subscale showing a mean rating higher than large class size was on the sensitivity of instructors to individual student and class ability to understand material presented in class (subscale #4). The small class size rating means were a full 1.5 on a five-point scale greater than the means for the large class size.

The amount of contact time with the instructor and the lack of opportunity to ask questions in large classes may have contributed to the discrepancy. When students can ask questions it tends to clarify misunderstandings and facilitates discussion that ultimately leads to learning. Input from a variety of perspectives is an interactional characteristic, and Wolff et al. (1987) argues that it stimulates discussion, hence learning in large classes. Conversely, if an instructor assumes understanding, and is unaware of student confusion or just does not care, the learning process is hindered.

These results were corroborated in the study by Aleamoni and Graham (1974) using multivariate analysis. The researchers investigated the relationship between student rating results (dependent variable) on six subscales of Course Evaluation Questionnaire instrument and student course level (independent variable). The purpose of their study was to determine if faculty members of higher ranks received the highest student ratings when class size and course level were taken into account in a sample of 448 course sections.

Multivariate analysis (MANOVA) results comparing class size, course level, and instructor rank indicated highly significant differences on rating scores at the course level, but no interaction was discovered for the class size variable compared singly by the six subscales between small (1-20), medium (21-40), and large (over 40) course sections or by academic rank from instructor to full professor (Aleamoni & Graham 1974).

Multivariate analysis (3 X 5 X 5 MANOVA) was also conducted in the present study comparing faculty rank, student class level, and class size (Minimum Class Size Policy). Interaction was discovered on all twelve instructional dimensions (SATI subscales) due to faculty rank and student class level, but not class size in the results of the variable analysis.

These results were statistically significant at the .05 level of confidence, but are of little practical significance in this investigation and most other studies. The magnitudes of the mean differences for faculty rank were small, less than 0.3 on most of the subscales; therefore caution should be exercised in interpreting the results on the faculty rank variable.

Hypothesis 10

The null hypothesis was unsupported, as significant differences were not found between student ratings of faculty teaching performance and faculty tenure status. An intuitive assumption made in academic communities is that the greater the instructional experience and age of professors the higher their teaching performance ratings will be and/or the more positive the correlation will be between student assessment of teaching performance and tenure status.

In the present study an attempt was made to correlate age of 645 instructors and the number of years teaching with overall faculty teaching performance ratings by academic rank, a low positive correlation of .02 resulted for all academic ranks. Other studies have concluded that age and instructional experience are also not significantly related to students' global assessment of faculty teaching performance (Clark, 1973; Elmore & Pohlmann, 1976, 1978; Marsh & Overall, 1979).

Some researchers found an inverse relationship between instructor age/teaching experience and global ratings (Linsky & Straus, 1975; Marsh, 1976; and Marsh, Overall, & Thomas, 1976); while others found a positive relationship (Delaney, 1976, 1977; Walker, 1969). Delaney (1977) reported a positive correlation of .05, and Walker a mean difference of .14 on a five-point scale between instructors below and above five years teaching experience. Differences were too small to be of any practical significance.

In one of the most extensive reviews of the research literature ever conducted (in excess of 300 research articles) on seniority, age of the instructor, and instructional experience of the college teacher Feldman (1983) concluded that these variables are largely unrelated. Those studies where a relationship was found between the variables just mentioned, Feldman noted it as being weak, or inversely related to the global evaluation subscale.

In this study he also summarized the results of 62 studies investigating academic rank and overall effective teaching results and found that a statistically significant association between these two variables did not exist (Feldman 1983). This was also evident in the present study. Additionally, the results from comparing academic rank and publication rates with faculty teaching performance results revealed a low negative correlation coefficient indicating that no systematic relationship exists between faculty publication rates and mean ratings of teaching performance for all levels of academic rank.

Hypothesis 11

The null hypothesis was supported, as significant differences were not found between student ratings of faculty overall teaching performance of non-international teaching assistants (NITA) and full-time faculty (F-TF). However, differences were found between these two groups at the instructional dimension level of analysis on several low inference SATI subscales.

The purpose of this aspect of the study was two-fold. First, to determine if students perceived the overall teaching performance of non-international teaching assistants (NITAs) and full-time faculty to be significantly different from each other and second, where did these differences, if any, occur along the 12 instructional dimensions of SATI?

It should not be surprising if non-international teaching assistants receive higher student ratings for overall teaching performance when compared to the international teaching assistants, the former being native born in the culture and language. But if NITA mean ratings on a specific rating dimension, i.e. communication skills subscale #10 of SATI, was significantly higher (4.2653) than those received by Associate (4.0562) and Full Professors (4.0041), then it is of some interest.

Newman-Keuls comparison of mean ratings of teaching effectiveness for the NITA an F-TF groups shows them to be statistically indistinguishable from the means of all academic ranks from instructor to full professor (Table #38). The average number of years teaching experience for the 645 full-time faculty in this study was 15.4 years as compared with 1.5 years of teaching experience for 193 NITAs. Faculty teaching experience ranged from three months to thirty year; NITAs ranged from three months to five years.

Theories of effective teaching suggest that the more classroom experience faculty have the better teachers they become, evidenced by significantly higher student ratings of teaching performance (Good &

Brophy, 1984). Assistant professors, based on the above premise, should receive lower ratings than a full professor and significantly higher ratings than a non-international teaching assistants would receive with only three months teaching experience, such is not the case in this study.

Results from a campus-wide analysis of 645 full-time faculty and 193 non-international teaching assistants suggest that it makes little difference what academic rank an instructor has attained or the number of years of classroom teaching experience. The rating means were statistically the same when the unit of analysis was academic rank irrespective of the department and/or college.

Feldman (1983) summarized the results of 62 investigations comparing levels of academic rank of instructors with their overall ratings of effective teaching and concluded the predominant outcome of the analysis was that a statistically significant association between these two variables was not found. Additionally, in studies where age of the instructor and instructional experience of the college teacher were correlated with academic rank and seniority, the overall results were similar, e.g. the variables were largely unrelated. Those studies that found a relationship were noted as being weak, or inversely related to the global evaluation subscale.

Feldman's study results concur with the results found in the present investigation. In the study by Centra and Creech (1976), no differences were found among the levels of academic rank from instructors to full professor, including teaching assistants; as were the results of the present study. Similar results were found in the investigations of Aleamoni and Brandenburg (1973) and Brandenburg, Slinde, and Batista (1977).

The investigation of Marsh (1976) also corroborate the above findings. He compared a group of teaching assistants with a group of full-time faculty; although undifferentiated by academic rank, the former group did received slightly lower ratings than the latter academic-rank conglomerate. Not all investigations reported teaching assistants with lower mean scores; Aleamoni (1972), Choy (1969), Morgenstern (1969), Nevill, Ware, and Smith (1978), and Weerts and Whitney (1975) found no difference in the global evaluation of full-time faculty and teaching assistants.

In summarizing the research results comparing NITA and F-FT groups on the overall or global subscales of various rating instruments, the results are similar to the results derived from comparing F-FT academic levels. The NITAs ratings are either slightly below F-TF or not significantly different from them.

A number of studies reported no association when academic rank was compared with the overall evaluation of teaching performance; however, some found one or more positive relationships between academic rank and certain specific instructional dimensions (Elmore & Pohlmann (1976); Linsky & Straus, 1975; Marsh, 1976; Marsh, 1980a; and Van Horn, 1968). Feldman (1983) provides convincing evidence why a more positive relationship is not found between academic rank and overall teaching performance when considering the enormous range in variables like age and number of years teaching experience among faculty. This can best be illustrated in the present study. How does a researcher compare teaching performance ratings of a 56 year old tenured assistant professor with 20 years of teaching experience in the College of Education with a tenured associate professor from the College of Science, age 33 with 7 years teaching experience, and derive meaningful results from the comparison?

The second part of the F-FT/NITA variable comparison was to compare all academic ranks on the basis of the 12 instructional dimensions of SATI. The results indicated the presence of significant difference between academic rank and students' perception of certain aspects of effective teaching among these ranks. Feldman (1983) noted from an extensive review of research on student evaluation of faculty teaching performance that overall rating results tend to mask significant differences between instructors on effective teaching variables that would otherwise allow differentiation among them on their ability to teach well.

Multivariate analysis (3 X 5 X 5 MANOVA) was conducted comparing faculty rank, student class level, and class size (Minimum Class Size Policy). Interaction was discovered on all twelve instructional dimensions (SATI subscales) due to faculty rank and student class level. Class size was not significant in the results of the variable analysis. Even though the results were statistically significant at the .05 level of confidence the magnitudes of the mean differences were actually quite small (less than .3 of a point) for faculty rank; therefore caution should be exercised in interpreting the results on this variable for most of the subscales.

The NITA (non-international teaching assistants) rank mean ratings were not significantly different from all other academic ranks on all SATI instructional dimensions, except for the use of good communications skills (#10). Students may have perceived the NITA instructor as being closer to their own age, making it easier to communicate more effectively at this level.

Assistant professors were perceived by students at all class levels to be better equipped in the areas of course organization/clarity (subscales #1, #2, #3), instructor/student rapport (subscale #4), stimulating enthusiasm for the subject matter in the courses taught (subscale #5), and on the learning a significant number of new ideas (subscale #11), as compared to all other academic ranks. The NITAs and instructors were rated by students as using significantly better communications skills compared to all other academic ranks (subscale #10); with the NITA rating means higher than instructors.

Both the assistant professor and the NITA ranks received the lowest evaluations means in the data analysis on examinations being relevant to reading assignments and materials presented in class (subscale #9); whereas associate and full professors were rated highest on this instructional dimension. They were also rated higher than other academic ranks on providing office hours for student consultations (subscale #6), dealing fairly with students (subscale #7), encouraging students to think independently (subscale #8), and on the overall subscale (#12), in which students indicated they were more favorably impressed by associate and full professors compared to other academic ranks.

The student class level variable results showed that freshman and graduate students gave the highest mean ratings on all 12 instructional dimensions, followed by seniors. This is consistent with the findings on hypothesis #9, where the mean ratings of freshman (100 level) and graduate (500 level) students were not only statistically different from each other and from the ratings of 200-400 level students, but significantly higher. These findings in the present study are contrary to the results obtained by Aleamoni and Graham (1974) within the course level variable analysis, which they also employed multivariate analysis.

Hypothesis 12

The null hypothesis was supported, as significant differences were found between student ratings of faculty teaching performance of non-international (NITA) and international (ITA) teaching assistants. The teaching performance of NITA and ITA groups were compared on the bases of student class status (freshman, sophomores, juniors, and seniors) and ratings they received on 12 SATI subscales.

The purpose of this aspect of the study was two-fold. First to determine how students at all undergraduate levels perceived the teaching performance of ITA and NITA groups along the 12 instructional dimensions of SATI. It is at the undergraduate level where the vast majority of student complaints about teaching assistant effectiveness originate and at in that level in which most of the research has been conducted with regard to TA teaching ability. Secondly, the study was intended to evaluate the effectiveness of TA preparation programs at O.S.U. based on composite rating scores TAs have received from their students.

Bailey (1984) coined the phrase, "the foreign T.A. problem," to represent problems encountered by undergraduate students in lecture courses, laboratory sections, recitations, discussion sessions, etc. taught by foreign graduate teaching assistants. The most significant problem was believed to be linguistic in nature, e.g. understanding subject matter and concepts filtered through a second language presenter.

Byrd and Constantinides (1988), Byrd (1987), Bernhardt (1987), and Rounds (1987) argued that English proficiency is an important basic skill; however, it does not constitute the communications problem. Adequate communicative competence for teaching assistants as a group involves the development and refinement of teaching skills, e.g. lecturing styles, appropriate teacher-student interaction, content organization, clarity of material presented, awareness of cross-cultural differences, and knowledge of expectations of students in American institutions of higher education.

Before any effective TA orientation program is established at a particular university or college for the purpose of improving undergraduate instruction, it is important to know first what the research literature has identified as problem areas and, second, if they occur at the institution under study. It is equally important that the investigation be comprehensive, e.g. across an entire university or throughout a college if possible, before the results can be generalized for the institution as a whole.

In the present study the teaching performance ratings of 192 teaching assistants (ITA and NITA) from nine colleges were analyzed. Interaction was discovered between the NITA and ITA groupings on five of the twelve organizational/clarity instructional dimensions (subscales) comprising SATI at a confidence level of .05; these were: #1 (Course objectives and requirements were clearly presented to me), #3 (The instructor explained the material clearly), #9 (The examinations were relevant to the reading assignments and to the material presented in class), #10 (The instructor used good communications skills), and #11 (As a result of having this instructor, I have learned a significant number of new ideas and/or skills).

Research results in the present study indicate that the NITA group had significantly higher overall teaching performance ratings and higher average means on all subscales of SATI when compared to the overall ratings discovered for the ITA group. These data are consistent with previous findings where foreign or international teaching assistants consistently received less favorable overall ratings by their students than their native or non-international teaching assistants counterparts. (Bailey, 1982, 1983, 1984; Barlow, (1985); Bernhardt, (1987); Byrd, (1987); Byrd & Constantinides, (1988); Rounds, (1987); and Shahenayati, (1987). The overall rating scores of both TA groups in the present study were above university standards, whereas in the others they were not.

Data on differences at the instructional dimensions level of analysis are lacking in the majority of other research investigations including those just mentioned. This is important to realize because in the present investigation, it was discovered that the overall rating value tends to mask significant difference that actually exist between groups compared on overall ratings and not along those instructional variables measuring effective teaching constituting the composite score.

Figure 4 shows the range in mean scores from low to high means at all course levels on five subscales given by ITA and NITA groupings where interaction was discovered. Student perceptions of the teaching effectiveness between ITA and NITA groups gradually widened on most instructional dimensions at each successive course level. This may be attributed to the accumulative contact time students had with their TAs at each class level; hence an increased refinement/sensitivity in discerning qualities of effective teaching an/or in formulating bias.

In general, the freshman level in the ITA grouping showed the least difference in mean scores on all subscales (0.4885) and seniors the most (1.9965) with the differences in mean ratings of sophomores (0.6125) and juniors (1.0232) levels falling in between. This is interesting from the standpoint that freshman are often portrayed as being the most sensitive overall to teaching inadequacies of international teaching assistants (Rounds, 1987).

The freshman (0.6593) in the NITA group appeared more sensitive to differences in the quality of instruction than to their ITA counterparts. This sensitivity diminished at the sophomore (0.3892) and junior (0.3506) levels then suddenly increased in magnitude at the senior level (1.7716) to nearly equivalent to what was found in the ITA grouping.

The difference in mean ratings between ITA and NITA groups can best be understood at the instructional level of analysis. Interaction was discovered between the NITA and ITA groupings on five of the twelve organizational/clarity instructional dimensions (subscales) of SATI, and main effects were found on the remaining seven subscales.

The first of these interactions was on subscale #1 (Course objectives/requirements were clearly presented). Freshman and sophomore class levels in the ITA group gave the highest ratings (3.8 points) on this subscale with a mean difference of only 0.2606 from the NITA group. Students at these levels seemed to know what to expect from their courses and also indicated with a mean rating of 3.6 that the ITAs had the ability to explain the course material clearly enough to be understood (subscale #3).

Sophomores and juniors also perceived ITAs as clearly conveying course expectations by rating TAs (greater than 3.75 points) who taught this group. However, senior students gave ITAs the second lowest rating of 2.6269 in the data set on this instructional dimension with the largest difference between the ITA and NITA groupings of 1.1159 on a five-point scale.

How clear course material was explained in class (subscale #3) was also perceived by students at the senior level to be as equally low in the ITA grouping; the rating means on this subscale were greater than 3.0 at all class levels. The ITAs teaching in freshman and sophomore classes received ratings greater than 3.5 and 3.9 in junior courses. These rating values exceeded what most senior faculty at the college level of analysis had received on these same subscales.

A main effect was noted on subscale #2 (The instructor was well prepared and organized) due to class level. Senior students also rated ITAs low on this instructional dimension with a mean of 3.2 to a mean rating of 4.0 for the NITAs. Freshman, sophomore, and junior level students gave ITA and NITA groups a mean rating of 4.0 on this subscale. The TAs who were perceived to be well prepared and organized were also perceived to be better presenters of both course material (subscale #3) and course objectives/requirements (subscale #1).

Lowman (1984) in his book, Mastering the Techniques of Teaching, state "The most prominent factors in student ratings of instructors concern clarity of presentation" (p.9). Subscale #3 was highly correlated .81 with instructional dimension #10 of SATI, which assessed student perceptions of the communications skills employed by TAs in general.

This includes linguistic ability (pronunciation being a main complaint with ITAs) and communications styles used to convey course concepts. More specifically, subscale #10 was meant to include conveying the nuances of the subject matter, articulation of ideas/concepts, and, importantly, the manner of presentation or information dissemination strategy employed. Students assigned more weight in terms of significantly higher mean ratings to the clarity of communication competence aspect of effective teaching than to student/TA rapport (subscales #6 and #7).

The mean difference between the ratings received by ITA and NITA groupings from freshman and sophomores students on communication ability was quite large, 0.7201 and 1.0112, respectively on a five-point scale. Although, students at all class levels gave the ITA group a mean rating of 3.35 on possessing good communications skills, the difference was particularly pronounced at these class levels.

Lower-division undergraduates are thought to be particularly sensitive to linguistic difference in their instructors (Bailey 1984). Exposure to linguistic peculiarities within an unfamiliar, competitive academic environment is thought to diminish the tolerance for linguistic/cultural diversity, while amplifying sensitivity to pedagogical differences (Rounds, 1987).

By the time ITAs teach the upper-division classes their communications skills have improved significantly. Senior students gave their highest ratings, relative to the ratings received on other subscales in this grouping, for the ability of ITAs to communicate effectively. The

ITAs teaching at the junior level received the highest mean ratings of 3.6116 of all class levels on communicative effectiveness. These particular TAs should have teaching responsibilities in lower-division courses where the mean difference between groups are the largest.

Why ITAs in the upper-division courses should be perceived by their students as being more effective communicators than at the lower-division level might be explained by examining the interaction between subject matter symbolism and ITA communicative competence. Upper-division courses may be more symbolism dependent by nature of the depth and breadth of the subject matter covered compared to the lower-division courses in a particular content area. Therefore, upper-division courses may require less linguistic competence to convey concepts that can be adequately expressed by formulas and equations; 97% of the ITAs in this study teach in the College of Science. This view is also held by Rounds (1985) and Wittrock (1974a, 1974b) who believed that, the more abstract science/mathematics concepts become or higher the course level in which they are presented, the easier it is for instructors to convey these concepts symbolically verses linguistically.

In the NITA grouping all of the high rating values were received on subscale #10, with one exception; freshman gave their highest rating for subscale #11 (I have learned a significant number of new ideas and skills from the instructor) as did the freshman in the ITA group. These results may be anticipated for the NITAs since they are born in the host language and culture, they reflect cultural nuances in which native born students can readily relate.

The <u>lowest</u> mean values for the NITAs at every class level, including junior and senior levels for the ITA group, were given on subscale #9 (Examinations were relevant to assignments and lecture). It was designed to measure the extent students perceived the relevance of assignments and graded materials in the course. On subscale #9 freshman gave a rating 3.4779, sophomores 3.7957, juniors 3.8054, and seniors 2.2938. A mean difference of 1.5 points over a five-point scale between juniors and seniors. The ITA results were the same on subscale #9, a sudden drop in student ratings from 2.9565 for juniors to 1.2487, a difference of 1.7 points. The difference in mean scores between a high of 3.8054 given to the NITAs by juniors to a low mean of 1.2487 given the ITAs by seniors was 2.55 on a five point-scale.

Clearly, low results indicate that students at all class levels for NITAs and junior/senior levels for ITAs perceived that reading assignments and content material presented in the course as not germane to examinations. Examinations/graded assignments do not adequately reflect course content as emphasized by the instructors. Seniors rated both the NITAs (2.2938) and ITAs (1.2487) well below university standards (less than 3.0 points) on this important effective teaching variable. Examinations given students by teaching assistants are often prepared by full-time faculty and/or reviewed for content by faculty if designed by TAs.

The results on subscale #9 may explain why senior students felt that they did not learn a significant number of new ideas/skills (subscale #11) in courses taught by ITAs. Examinations should reflect those salient aspects of course content; when they do not students' receive mixed signals as to what instructors consider relevant in their courses.

Students at the freshman, sophomore, and junior class-levels perceived ITA instruction as having availed them with a significant number of new ideas and skills. The mean ratings were greater than 3.5 points, with junior- level students rating the ITAs as high as the NITAs on this instructional dimension. The trend in mean scores increased from 3.4653 for freshman to 3.5504 with sophomores to a mean of 3.9797 at the junior level, then suddenly drop to a mean of 2.8827 for the Seniors; a class-level difference greater than 1.0 between junior and senior students. The same was discovered for the NITAs, although, the sudden decrease between junior and senior levels was less than 0.2 of a point.

Main effects were found on seven of the twelve SATI subscales. Effects were due to influences other than TA or class levels on subscales #2, #6, #8, and #12 and to the TA on subscales #4, #5, and #7. Subscale #2 is an organizational/clarity instructional dimension and #12 an overall or global subscale. The remaining five subscales were classified as belonging to a student/instructor rapport (#4, #6, #7, and #8) and enthusiasm (#5) instructional dimensions.

The student/instructor rapport instructional dimension is considered a low inference variable influencing ratings of effective teaching. It is a variable that note changes in student perceptions over time, although not enough to change the overall perception. High inference variables are ones in which students make quick and long-lasting first impressions about instructor abilities/personality characteristics, which would include communications skills, knowledge of subject matter, clarity of presentation, instructor warmth, and concern for student learning (Smock & Crooks, 1973).

Results on subscales #4 (The instructor was sensitive to my/the class' ability to understand the material) and #7 (The instructor was fair and impartial in dealing with me) of SATI showed that students in both the ITA and NITA groupings at all class levels percieved instruction as being equally effective. Both groups gave faculty ratings greater than 3.8 points except for seniors who gave a mean of 3.5 to the ITAs and 4.0 to the NITAs on both subscales. In contrast, students at all class levels, especially freshmen and seniors, perceived the ITAs having stimulated significantly less enthusiasm in their classes (subscale #5), e.g. displaying less energy, humor, and ability to hold student interest with the presentation style employed than their NITA counterparts.

These results are consistent with the findings of Barlow (1985) who found that students' perceptions of TA enthusiasm significantly differed between the beginning and end of the semester when the TAs were provided feedback by trained observers to improve on this and similar instructional dimensions. The most pronounced difference in this study was on the Stimulation of thinking variable. The variable equivalent on SATI is subscale #8 (The instructor encouraged me to think for myself).

In both studies TAs that did not encourage class discussion did not invite students to share their own ideas and/or allowed students to be critical (constructive) of those ideas presented by the TAs were rated significant lower than TAs that encouraged these types of behaviors. In the present study the NITAs received mean ratings greater than 0.5 higher than the ITAs at all class levels on this instructional dimension.

The importance of stimulating and encouraging students to think can be seen in the investigation of Shirvani (1987) who compared differences between communication styles and teaching effective of native and non-native teaching fellows. Native student perceived non-native teaching assistants as being less relaxed (not calm but nervous under pressure), less encouraging, less friendly, more contentious, and generally poorer communicators, compared to native TAs.

The investigator attributed student perceptions of ITAs to unfamiliarity of American pedagogical methods, weak linguistics skills, cultural differences, and a lack of classroom management skills. The ITAs were also thought of as being more conservative and less dramatic than their NTA counterparts.

Subscale #6 (The instructor provided scheduled office hours or was readily available for consultation with me) was the last instructional dimension on which a main effect was found. Lower-division students perceived both NITA and ITA groups as being equally effective in providing one-on-one consultation time by having given a mean ratings greater than 3.8. At the upper-division level ITAs teaching juniors were rated 0.6 lower than the NITAs. Senior students perceived both the NITAs and ITAs as not having provided the necessary out-of-class student/instructor scheduled contact time and/or being available for one-on-one consultation.

The second part of the analysis was designed to compare the TA orientation seminar and departmental level training programs implemented at O.S.U. for the purpose of improving undergraduate instruction. Both forms of training were reported to have focused on certain aspects of the linguistic, pedagogical, and cross-cultural communication components of communicative competence that would facilitate ITA teaching ability and adaptation to the university's instructional environment.

The results indicated that the TA preparatory training at the level of departments was significantly more effective than the ITA Orientation Seminar offered by the Postsecondary and Technological Education Department in the College of Education. In fact, the ITAs in the College of Education received the lowest overall ratings of teaching effectiveness; whereas the ITAs in the Colleges of Science and Engineering received the highest ratings.

The TAs who received both departmental training and attended the ITA Orientation Seminar received mean ratings that were not statistically different from those TAs who just attended the ITA Orientation Seminar, despite the double exposure to effective teaching/classroom management training. Although, TAs who did attend one or the other training programs, or both, did significantly better overall than TAs who had no formal training prior to being assigned teaching responsibilities.

The ITAs who had gone through preparatory training at the department level may be perceived by students as being more effective teachers because of familiarity with the subject matter and teaching styles endemic to their discipline. Students come to expect certain pedagogical approaches that allow them to receive and process information in a preferred learning style compatible with the nuances of a particular discipline. The TAs are products and models of academic disciplines.

Generic TA training programs provide only generic training on effective teaching variables. While these variables do cut across most disciplines, effectiveness may vary as a function of the type of teaching style employed (Hudak & Anderson, 1984).

In the findings of Basow and Distenfeld (1985) and Basow and Silberg (1987) students majoring in engineering gave professors in the humanities and social sciences the least positive ratings on each instructional dimension on the rating instrument used in the study. These researchers argued that engineering students may have been less accustomed to teaching styles in the areas of social science and humanities or may have had less interest in these courses than other students.

This unfamiliarity with teaching styles in tandem with a competitive academic environment characterized by linguistic/cultural diversity in the instructional staff may have contributed to what Rounds (1987) observed as the reduction in tolerance levels of native students towards foreign TAs. This intolerance for ITAs surfaced as complaints of extreme pedagogical differences that native students perceived as encumbering the teaching/learning environment.

In summary, the results of the present study indicate that undergraduate students perceived the teaching effectiveness of international and non-international teaching assistants differently. These perceptual differ-ences ranged along class level on a number of instructional dimensions.

Senior-level students were found to be the most sensitive to variations in pedagogical abilities among ITA and NITA groupings. They consistently rated ITAs below university standards on clearly presenting course objectives/requirements, relevancy of examinations to assignments, and having learned a significant number of new ideas from the instructor in the course. The NITAs were also rated below university standards on the examinations instructional dimension.

Freshmen and sophomores were found to be sensitive to communications abilities of the ITAs. This may be attributed to unfamiliarity with individuals from different linguistic and cultural backgrounds. However, upper-division students rated ITAs as being competent communicators.

This may be the result of synergy between subject matter symbolism and ITA communicative competence but perhaps more to improvement in linguistic competence with classroom experience and cultural immersion.

The ITAs and NITAs were perceived by their students as fair in interacting with students and concerned about the ability of the individuals and/or the class' ability to grasp subject matter concepts. Only seniors were negatively affected by this instructional dimension or the lack of it in ITA teaching behavior. Even though ITAs did receive a mean rating of 3.5 on this subscale, which is above university standards (above 3.0 on SATI), special attention should be focused on improving ITA classroom behavior that would change student perceptions that these needs are being met.

Another area that ITA training should focus on is the enthusiasm they show for the subject matter and discipline itself. Students at all class levels, especially freshman and seniors, perceived the ITAs as greatly lacking enthusiasm for the subject matter taught in the course. This may be due to a more traditional (less energy, humor, and ability to hold student interest) approach to teaching than lack of genuine concern. Students perceived the ITAs as being well prepared and organized (subscale #2) but lacking in enthusiasm.

Lower-division students rated both the NITA and ITA groups high, a mean ratings greater than 3.8 points, or equally effective in being available for after class consultation. The ITAs teaching upper division courses were rated 0.6 lower than the NITAs, especially senior students who rated both the ITA and NITA groups low on this instructional dimension.

The second part of the analysis focused on the ITA teaching preparation for classroom duty. Departmental-level training programs are the more effective in preparing ITAs as perceived by the students taught by these teaching assistants. The ITA Orientation Seminar offered by the Postsecondary and Technology Education Department was perceived by students as being less effective but more effective than no teacher preparatory training at all. Training in an artificial verses an actual classroom may be the reason. When teaching effective was computed for ITAs at the college-level unit of analysis the College of Education received significantly lower overall ratings (3.1) as compared to the Colleges of Science (3.5618) and Engineering (3.6131).

Hypothesis 13

The null hypothesis was supported, as significant differences were not found between faculty publication rates and student ratings of faculty teaching performance. In this study faculty productivity was operationalized as scholarly productivity and defined as the number of articles published by faculty in refereed journals over a period of five years. The 645 faulty who participated in this study had instructional responsibilities of no less than six credit hours per academic year.

The purpose for this phase of the investigation was to determine if there was a relationship between one dimension of faculty productivity, i.e. publication in refereed journals, and faculty teaching performance ratings. Specific questions that this aspect of the study focused on were:

1. Is there a correlation between faculty publication rates and faculty teaching performance ratings?
2. Are there differences in productivity levels associated with various academic ranks?
3. Does productivity vary with tenure status?
4. Are publication rates, before tenure review, associated with later publication rates?
5. Is there a difference in publication rate and faculty gender?
6. Are there differences in levels of productivity between colleges?
7. Does the productivity level of the College of Education at Oregon State University vary from colleges of education in other universities in the Pacific Athletic Conference?

Faculty publication rate data meeting TALAGLIS criteria were averaged over four academic ranks: instructor, assistant professor, associate professor, and full professor. Publication rate means were then correlated with teaching performance means for each academic rank and for all ranks combined (Table 40). The results are discussed below with their respective question. Before discussing the data supporting each question raised in hypothesis 13, a brief discussion of the importance of publishing to an institution, faculty, and students is in order.

The most commonly accepted mode of scholarly exchange of ideas between faculty within an academic community and the most commonly accepted means of maintaining currency in a discipline are through the publication of articles in journals that ascribe to the highest standards of research in a particular discipline. Quality publishing involves conducting quality research. The reputation of a comprehensive research-oriented institution is strongly correlated with the research productivity of its faculty (Blau, 1973). This is equally true of the academic units subsumed within the institution. Aleamoni (1987) noted a substantial increase in the use of publication rates in professional journals from the year 1978 to 1983 in the evaluation of the research component at most universities. Prior to 1978 the emphasis was on authoring and editing books, writing monographs, and editing books and not professional journals.

The tripartite role prescribes that faculty wishing to be tenured or promoted to the next academic rank at Oregon State University must publish. Most department heads and college deans are aware of the influence that faculty publishing has on securing federal, state, and local research funds and the influence it has on graduate student research and their later employment opportunities.

Carleson (1985) compared the publication history of faculty and their graduate students, a correlation of .89 was obtained. He concluded that faculty who publish had students who would also publish throughout their careers. Conversely, faculty who did not publish had students that focused almost exclusively on other aspects of the tripartite role to the exclusion of research and publishing. At the university in which this research was conducted hiring practices included an investigation of the publication history of major professors of those candidates applying for positions.

There is a high correlation of .87 in the present study between those faculty who wrote grants and trends towards increased productivity at successive promotion levels. The university receives approximately 130 million dollars annually in research grants compared to the 30 million dollars for the University of Oregon and 15 million dollars for Portland State University.

Contrast this with an average teaching load of from 9 to 15 hours for faculty at O.S.U, 7 to 9 hours at P.S.U, and 3 to 5 hours U.of O. Faculty at O.S.U. have a substantially higher teaching load, yet manage to acquire four to six times more grant monies through writing and publishing efforts than all institutions in Oregon's higher education conglomerate (Oregon State System of Higher Education, Chancellors Office, Sept. 27, 1989).

There are seven sub-questions to hypothesis #13. All are found in previous research that have compared faculty productivity and teaching performance ratings and therefore should be included in any serious study of the relationship of these two variables. Selected individual characteristics, which make up the sub-questions, include the following variables: academic rank, tenure status, academic unit of employment, number of publications at time of tenure/promotion review, and gender of the instructor.

In general, the grand mean of scholarly productivity was 4.8879 (n= 645), or approximately one article published per full-time faculty member per year for all academic ranks. The main effect of gender on productivity for each of the selected individual characteristics were not significant. However, variations in productivity by gender were found and will be discussed.

In the Huck, Cormier, and Rounds (1973) measure of central tendency and variability a correlation coefficient like the one just described above of .87 between those faculty who wrote grants and increased productivity at successive promotion levels would be considered to be a high positive correlation (Refer back to hypothesis 13 in Chapter Four of this compilation for further explanations of correlation coefficient magnitudes).

<u>Question 1</u>: Is there a correlation between faculty publication rates and faculty teaching performance ratings?

Pearson product-moment correlation results yielded a correlation of -.0293 for faculty ranks overall, corresponding to p= .05 value. Low overall coefficients indicate that no systematic relationship exists between faculty publication rates and mean ratings of teaching performance for all academic ranks.

In a meta-analysis conducted by Feldman (1987) comparing faculty teaching performance ratings and productivity rates, found that "research productivity is positively but very weakly correlated with overall teaching effectiveness as assessed by students."
(p. 240). He concluded by stating, "That research at least does not detract from teaching and might even have a slight likelihood of benefiting it may indeed be true, but this conclusion does not necessarily follow from the observed positive correlations that have been found, for its is conceivable that relatively effective teachers who are also productive in research would be even more effective were they to do less research" (p. 246).

It made little difference in Feldman's investigation as to what variable(s) was used to define productivity, the association was always weak between these two variables. This was also the result when studies were comparing faculty productivity and student rating results along individual instructional dimensions. Low association between evaluation of teaching effectiveness and research productivity was found to be generally very small on both the overall score and on each instructional dimension.

<u>Question 2</u>: Are there differences in productivity levels associated with various academic ranks?

An ANOVA was run to determine whether faculty in various academic ranks showed significant differences in productivity. Faculty were grouped as instructors, assistant professors, associate professors, and full professors. A large, significant difference was found for productivity among the four groups. Newman-Keuls revealed that the overall productivity means of associate and full professors were statistically indistinguishable. Means for instructor and assistant professor where distinguishable and quite large by comparison.

A negative relationship resulted between course evaluations and publication rates overall. The correlation coefficients for assistant professors was -.0248 and associate professors it was -.0297. Instructors showed a positive correlation of .0487 as did full professors (.0099). Low overall coefficients indicate that no systematic relationship exists between faculty publication rates and mean ratings of teaching performance for all academic ranks.

A high correlation of .73 was found between productivity and increasing rank for all faculty. Fulton and Trow (1974) found increasing productivity with higher ranks, and Finkelstein (1984) concluded that, even when the incentive of promotion was no longer important to faculty in higher ranks, continued productivity "supports, once again, the preeminent role of 'intrinsic' as opposed to 'extrinsic' motives for research involvement" (p. 101).

Question 3: Does productivity vary with tenure status?

An ANOVA was run to determine whether faculty showed significant differences in productivity with tenure status. Differences were found for productivity among the three tenure groups. Tenure level designated as indefinite tenure were significantly higher than fixed and annual statuses. Annual tenure productivity was significantly higher than what was noted for the fixed-tenure status.

When Pearson product moment correlation coefficients were computed for each tenure level comparing publication and teaching performance means the results indicated a low negative relationship existed between these variables. The ANOVA results comparing publication rates, teaching performance means and tenure status also showed no statistically significant difference between these variables.

However, ANOVA results comparing publication rates and tenure status alone revealed a statistically significant differences, as well as a high Pearson product moment correlation value of .6421 for all tenure levels combined, corresponding to p= .05. For the fixed term tenure status correlation coeifficents was equal to 0.6373, annual .6296, and indefinite .7162. High positive correlations indicate a systematic relationship exists between faculty tenure status and publication means.

Question 4: Are publication rates, before tenure review, associated with later publication rates?

An ANOVA was used to determine whether tenured faculty with different levels of early productivity showed significant differences in later productivity. Faculty were grouped according to the number of publications at the time of tenure review: 0-2; 3-5; or more than 6 articles published. A statistically significant difference for present productivity was found among the three groups. Publications at the time of tenure review were correlated with later productivity for all faculty.

The resulting correlation coefficient of .78 indicates a very strong relationship between faculty publication histories before and after tenure was awarded. This was especially noticeable in the data set at the college level unit of analysis. These data indicate that faculty with low productivity levels before tenure continued to have low productivity levels later in their careers (mean of 1.6), e.g. Colleges of Education and Home Economics. Conversely, faculty who showed high productivity levels early in their careers continued to have high levels of productivity later on (mean of 9.9), e.g. Colleges of Agriculture, Science, and Engineering.

<u>Question 5</u>: Is there a difference in publication rate and faculty gender?

An ANOVA was used to determine whether female faculty differed from male faculty in productivity rates. Results showed significant differences between these two groups. Male faculty published on the order of two articles to female faculty's one across all colleges. Additionally, female faculty productivity tended to drop off with each academic rank attained after tenure was awarded; whereas for male faculty the trend was to increase in levels of productivity each academic rank awarded with a slight decline after full professorship was awarded. The correlation between productivity and gender was a positive .63 for male professors. The correlation for females was both negative and nonsignificant (-.07). Female faculty also showed no correlation between tenure and productivity.

<u>Question 6</u>: Are there differences in levels of productivity between colleges?

The range in productivity levels between colleges was significant. The College of Education showed the lowest level of productivity (mean of .5714) and the College of Pharmacy had the highest productivity (mean of 13.0). The overall mean for all colleges combined was 4.8879 articles published over a five-year period. For the university one article published per year was the overall publication index for all colleges combined.

At the college level the relationship "r" was even higher for the Colleges of Science and Agriculture between productivity and increasing rank (.89). Full professors surpassed all other academic ranks in their productivity among the three groups. They also had a heavy teaching load. Interestingly, the average class size for full professors tended to be larger than the mean for all class sizes combined and for all other academic ranks, a mean of 29 students per class as compared to a mean of 23.0 students overall. It would seem that seniority would have its privileges of smaller class sizes, but in this case full professors also had full class sizes and full teaching loads.

In the Colleges of Education, Health/P.E., Home Economics, Engineering, and Pharmacy there was a sudden drop-off of productivity at the full professor rank; only the Colleges of Science, Agriculture, and Liberal Arts maintained a steady increase in productivity rates at this rank. The productivity rates of Education, Liberal Arts, and Business were lower than the overall productivity rate for all academic ranks and colleges combined. When tenure was awarded at the rank of assistant professor productivity decreased by half for Health/PE, while it doubled in the Colleges of Education, Home Economics, and Business for associate professors which was their most productive rank.

At the academic rank level of analysis in the colleges mentioned above, only the productivity rate of assistant professors in the Colleges of Science, Agriculture, Engineering, and Health/P.E. exceeded the overall productivity rate for assistant professors combined. Only the productivity rates in the College of Education at all academic ranks were far below the overall productivity rate computed for the entire university.

Carleson (1985) compared the publication histories of faculty and their graduate students and found that faculty who frequently published had graduate students who also published regularly in their career disciplines. The author also noted that faculty who publish frequently early in their careers tended to publish regularly later in their careers.

To test Carleson's second finding productivity rates of faculty in the College of Education were traced to 1961, 30 years. Productivity indices were computed for each faculty before tenure was awarded and at each successive promotion level. The results show that tenure and promotion was awarded without meeting the university publication index of one article per year, or the indices computed by academic ranks. Comparisons with other colleges of education confirms the low productivity rates of O.S.U.'s College of Education, although this condition has its antecedents in the past.

The college level of analysis may be too large for a comparison of academic units, if that were even possible with the enormous variation in goals, objectives, and missions between them. However, departments within the same college may be more homogeneous in terms of their mission, hence comparisons more reasonable.

Some departments within colleges are extremely productive, while others are not. The efforts of those productive departments are "blended" out at the college level of analysis by those that choose to fall behind university standards. For example, the overall publication rate for the College of Liberal Arts was computed to be 1.9635 articles over a five year period. However, the Department of English produced a publication rate of 4.6831 articles in referred journals over this same time period. This rate is substantially higher at the department level than departments combined within the college. The present research was restricted to the college as the unit of analysis by agreement with the university in allowing this research to continue.

Publication histories for colleges are based on faculty publication initiatives established over a period of decades. The current publication history of the College of Education at O.S.U., as in other colleges, did not suddenly emerge. When traced back 30 years a publication trend of one article per faculty member over a five year period was also found.

Question 7: How does the productivity level of the College of Education at Oregon State University compare with colleges of education in other universities in the Pacific Athletic Conference?

The College of Education at Oregon State University showed the lowest level of productivity (mean of .5714) among all colleges of education randomly selected from universities in the Pacific Athletic Conference (PAC 10). In fact, the lowest overall productivity rate found among the five colleges of education in the PAC 10 was eight times greater than the productivity rate computed for the College of Education at O.S.U. Even at the college level of analysis at O.S.U., the productivity rate of the college with the lowest value was three times greater than found for the College of Education.

The overall publication mean for all academic ranks combined was 5.9069 articles published over a five year period. Comparing academic ranks, instructors published least of all. Although, instructor publication (productivity) rates were higher in all PAC 10 colleges of education than at all academic ranks in the College of Education at Oregon State University. This deficiency was just as evident when O.S.U was compared with other universities on the basis of being a land grant university with similar demographics, e.g. W.S.U. and A.S.U.

The trend was to become more productive with increased academic rank at Stan. U., U.S.C, and U.C.L.A. The reverse was noted for O.S.U, W.S.U, and A.S.U., with a gradual decrease in publication rate to the rank of full professor. Full professors at O.S.U. stopped publishing altogether when this rank was attained.

Complaints from faculty in the College of Education at Oregon State University as to why they do not conduct research and publish their findings fits three general categories: (1) they do not know how to conduct research and publish results, (2) there are insufficient funds to conduct bona fide research, and (3) there are too few professional education journals in which to publish research results. Faculty cannot be totally blamed for the current condition of low productivity. Results from the present study show that the majority of faculty in the College of Education, historically, have been tenured and promoted without meeting the tripartite component of research and publication. Where is the impetus to produce scholarly work, regardless of funding or ability to do research, when tenure and promotion rewards are granted to faculty without their meeting this responsibility? The whole notion of faculty renewal in higher education across America is premised on the belief of developing and rewarding faculty creativity.

The first of the three complaints mentioned above is incomprehensible from the standpoint of faculty being hired at a research-oriented institution and having conducted research in their own Ph.D. programs. What does this suggest in terms of the quality of research of graduate students guided by faculty who argue that they do not know how to conduct research, especially if the complaint has some basis in fact? It suggests a low quality student research product historically.

The second complaint is more plausible than the first; when the effects of Measure Five are considered and the dramatic effects its had on the availability of general operating funds, the validity of the argument increases. These program reductions in the College of Education have been taking place gradually over the past five years, not just suddenly with Measure Five. During this period, as now, a number of "projects" were underway in the College of Education. If conducted competently these projects would have yielded results worthy of publication in refereed journals, yet few results ever made it into print to be shared with other professionals.

The third complaint was the paucity of professional journals in which to publish research findings. There is no legitimate excuse why faculty cannot publish in referred journals in education. There are in excess of 100 professional journals in education in O.S.U.'s Kerr Library alone that faculty in their respective disciplines can submit articles for publication. These journals are reproduced in Appendix B of this study.

Publishing is a shared commitment by faculty, college administration, and central administration. Priorities must be established and implemented in the form of goals and objectives a college would like to achieve (mission statement). A balance is needed in the College of Education at O.S.U., one that reflects faculty renewal, cooperative research with other colleges on campus, and most important of all a sincere desire to communicate via research literature with faculty in other colleges of education across the nation via a commitment to conduct research and publish.

Suggestions for Further Research

This study has determined that, in general, select demographic variables do influence student ratings of faculty teaching performance depending on the variable and how it is analyzed. However, many questions remain to be answered, and some additional issues have been raised in the course of this study. First, the literature reports that female faculty, especially in non-traditional disciples, tend to receive less favorable overall ratings due to student gender-stereotypic behavior and attitudes that are reflected in their evaluations of faculty teaching performance.

The present study poses an intriguing problem in that findings suggest that the reverse is true at O.S.U.; female faculty received higher overall ratings than their male faculty counterparts. This was particularly true in the non-traditional discipline of Engineering. Male faculty, while rated low in Engineering, received significantly higher ratings in Home Economics than did female faculty. These rating differences whether due to faculty gender/classroom behaviors or student gender-stereotypes/biases, should nevertheless be explored.

Student evaluation results continue to play a key role in faculty tenure and promotion decisions within universities across America. As such, an investigation designed to generate empirical evidence on what effects demographic variables might have on the effective teaching ability of faculty is a logical first step. The next step is to develop a predictive model that explores the interactions/interrelationships of all variables together as would occur in a real classroom situation, such that faculty teaching performance rating indices could be determined and analyzed in summative and formative evaluation processes.

More importantly, it would add a quantitative dimension in the assessment of just how effective faculty teaching was and on what instructional dimensions. The present investigation showed that overall rating scores tend to mask teaching deficiencies on instructional dimensions that impede student learning.

The Colleges of Engineering, Pharmacy, and Science are known to be academically rigorous. Just how prevalent this perception is among students, and how much it influences student ratings of faculty teaching performance should be carefully investigated; especially, in terms of the interaction between preferred learning styles of students and teaching styles of faculty.

These same three colleges were found to have the highest productivity rates compared to other colleges in the study but the lowest ratings of effective teaching at the undergraduate level variable analysis. A correlation of -.0293 was found between faculty publication rates with faculty teaching performance ratings for all academic ranks, indicating that no systematic relationship exists between faculty publication rates and mean ratings of teaching performance for all academic ranks. Research may be content dependent, something that students are not competent to judge; whereas effective teaching may be dependent on content organization, instructional delivery, and classroom management which students are competent to evaluate. The lack of correlation between the domain of publication and effective teaching may be too great for a meaningful comparison.

Why some colleges do better than others in training ITAs for classroom instruction should be thoroughly investigated across campus. It would be of equal importance, as a starting point, to conduct a survey of a large number of students at all class levels and in all colleges across campus about ITA instructional effectiveness. These data could then be compared with actual student ratings of ITA teaching performance. The blend of qualitative (which would include classroom observations) and quantitative data would provide foundational data to develop or improve upon existing ITA training programs at the department level.

The results of comparing student opinions with actual rating results may be similar to research on class size. Students (and faculty) have indicated the preference for small classes, yet no significant difference was found between faculty rating results in large or small class sizes in the present study.

Recommendations to Central Administration

Recommendations that have practical significance in addressing problems must first be based on facts related to the problem, and second possess sufficient specificity in the interpretation of gathered facts and based on the research literature, to function as change elements. This researcher feels that the following five recommendations contains both of these essential components.

Recommendation 1: Eliminate the ITA Orientation Seminar and reallocate funds to the department level of ITA training. Results from the present study show that the most effective TA training program occurs at the departmental level. Over the years much ado has been made about the "foreign TA problem" at O.S.U., and thousands of dollars spent on the international teaching assistant (ITA) Orientation Seminar sponsored by the Department of Postsecondary and Technological Education Dept. The purpose, as stated in O.S.U.'s Strategic Plan, was to establish guidelines and evaluation procedures for training foreign TAs to meet the demands of the instructional environment at the university.

Collette and Chiapptta (1986) argued that evaluation is fundamental to the improvement of instructor performance and course effectiveness by drawing attention to key instructional areas in the assessment procedure and correcting deficiencies that are found. Yet historically, the directors of the ITA Orientation Seminar have not made an attempt to evaluate the instructional effectiveness of their program where the research literature on effective teaching says it counts the most, e.g. based on student perceptions (responses gathered on inventories like SATI) of instructor teaching performance in actual classroom situations.

This is why departmental-level ITA training is more effective than the ITA Orientation Seminar. The former improves upon ITA instructional abilities by identifying strengths and weaknesses in actual classroom situations and at the appropriate student level while the later relies on the hypothetical classroom where the opinions of other ITAs acting as students form the basis for instructional orientation and/or improvement. Most of these ITAs lack prior pedagogical training and teaching experience in general and in American institutions of higher education in particular.

Recommendation 2: Course examinations must be made congruent with course content. Undergraduate student responses on SATI (subscale #9) in every college and department across campus noted this instructional dimension to be the achilles heel in University instruction at the lower academic ranks. This was especially evident at the teaching assistant (both ITA and NITA), instructor, and assistant professor ranks. Content expertise does not automatically confer the ability to teach or how to effectively evaluate what was taught.

The problem in the above ranks is not subject matter incompetence nor is it coming to lecture and laboratory unprepared or disorganized. It is, however, a perceptual difference between instructors and students about what was or what was not covered on examinations from reading assignments and material covered in class. The TA training should pay particular attention to this instructional aspect. Assistant professors have to gain this knowledge through classroom experience alone since a formal formative evaluation system does not exist at O.S.U.

<u>Recommendation 3</u>: A formative evaluation system is needed at O.S.U. in conjunction with the summative evaluation structure that is already in place. Student evaluation results show that not only certain academic ranks have fallen below university standards on certain instructional dimensions but colleges as well.

If the number one goal of O.S.U.'s Strategic Plan is to elevate and reward teaching on par with the research component of the tripartite role of faculty, then means must be made available to raise the level of instructional quality. This is only possible when formative and summative evaluation processes are in place that can improve instruction; unless, of course, O.S.U. hired only new instructional staff from other American universities that have already demonstrated an ability to teach. But this still would not correct the existing problem.

Summative evaluation processes locate sources of instructional deficiencies and may even identify what they are, but only formative evaluation input can provide means to correct them. It is not uncommon for faculty, even senior faculty, to be deficient in one or more instructional dimensions that reduces teaching effectiveness; hence student learning/course satisfaction.

Research shows that if not corrected these same problems tend to reoccur year after year in the same faculty. Even when faculty are shown student rating results, they often cannot translate these data into classroom teaching behaviors that would result in improved instruction. Formative evaluation strategies would provide expert consultation concurrent with interpretation of student rating results to produce the desired instructional outcomes. The results from the present study provides the crucial data necessary to improve university instruction, and the bases for creating a formative evaluation system in tandem with the existing summative evaluation system.

Recommendation 4: Tenure and promotion guidelines must be implemented consistently across all faculty and colleges. Findings in the present study show that faculty in the College of Education have not, past or present, met the tripartite responsibility of research and publication. A fundamental conflict may exists between faculty desires and central administration expectations. Most faculty in Education have expressed a desire to teach only, and student rating results from this study show that they do it well, having received among the highest ratings when compared along all demographic variable analysis and at the college level of analysis. However, their productivity rates are dismal by comparison.

If any conclusion can be arrived at from the present study it is the need to partition Education faculty FTE into the three areas of responsibility to encourage and reward progress in conducting research and publication. Also needed is an established productivity benchmark for faculty to attain.

Recommendation 5: Publication benchmarks are needed at the level of academic ranks and colleges. In the present study these benchmarks are provided. Individuals joining the faculty in a particular college or already on the faculty and wishing to be tenured and/or promoted now have a productivity-rate reference point. This point is based on the productivity rates of faculty in a college over a period of years. In situations where colleges are so far adrift, as in the case of the College of Education, reference markers from other universities are needed to establish a new bearing, which initiates movement of college faculty out of the doldrums into currents of contemporary research.

REFERENCES

AAHE Task Force on Professional Growth (1984). " Vitality without mobility: The faculty opportunities audit." In <u>Current Issues in Higher Education</u>, Series 4. Washington, D.C.: American Association for Higher Education.

Abaneme, M. K. C. (1987). The influence of motivation and achievement on student ratings of instruction: A comparative study of full-time an part-time teaching effectiveness. <u>Dissertation Abstract International</u>, <u>49</u>, 5A. (88-12,410)

Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). Educational seduction. <u>Review of Educational Research</u>, <u>52</u>(3), 446-464.

Abrami, P. C., & Mizener, D. A. (1985). Student/instructor attitude similarity, student ratings, and course performance. <u>Journal of Educational Psychology</u>, <u>77</u>, 693-702.

Alciatore, B. T., & Eckert, R. E. (1972). A college teaching internship program. <u>Improving College and University Teaching</u>, <u>20</u>, 93-96.

Aleamoni, L. M. (1972). <u>The Illinois course evaluation questionnaire: Manual of interpretation</u>. (Research Rep. No, 331). Urbana-Champaign, Ill.: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Aleamoni, L. M. (1976). Typical faculty concerns about student evaluation of instruction. <u>National Association of Colleges and Teachers of Agriculture Journal</u>, <u>20</u>(1), 16-21.

Aleamoni, L. M. (1978). The usefulness of student evaluations in improving college teaching. <u>Instructional Science</u>, <u>7</u>, 95-105.

Aleamoni, L. M. (1986). Developing a comprehensive system to improve and reward instructional effectiveness. <u>Resources in Education</u>, <u>19</u>(1), 45-32.

Aleamoni, L. M. (1987). "Techniques for evaluating and improving instruction." New Directions for Teaching and Learning, no. 31, San Francisco: Jossey-Bass.

Aleamoni, L. M., & Brandenburg, D. C. (1973). Optional item catalog (Research Memorandum No. 146). Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois at Urbana-Champaign.

Aleamoni, L. M., & Graham, M. H. (1974). The relationship between CEQ ratings and instructor's rank, class size and course level. Journal of Educational Measurement, 11(2), 189-204.

Aleamoni, L. M., & Hexner, P. Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. Instructional Science, 9, 67-84.

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

Aleamoni, L. M., & Yimer, M. (1973). An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. Journal of Educational Psychology, 64, 274-277.

Aleamoni, L. M., & Yimer, M. (1974). Graduating senior ratings' relationship to colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. (Research Report No. 32). Urbana: Office of Instructional Resources, Measurement and Research Division, University of Illinois.

American Association of University Professors (1915). Academic Freedem. Fall Bulletin, p. 132.

Anukwuem, C. U. (1987). Student rating of instructors as related to instructor demographic variables. Dissertation Abstracts International, 46, 12A. (University Microfilms No. 86-03,749)

Ard, J. (1987). The foreign TA problem from an acquisition theoretic point of view. English for Specific Purposes, 6(2), 133-144.

Armenti, Jr. A., & Wheeler, G. F. (1978). Hawthorne effect and quality teaching: Training graduate teaching assistants to teach. American Journal of Physics, 46(2), 121-124.

Arnst, D. (1978). Improving instruction: Reform the institution, not the faculty. Liberal Education, 64, 266-277.

Arreola, R. A. (1979). Strategy for developing a comprehensive faculty evaluation system. Engineering Education, 14, 239-244.

Arreola, R. A. (1984). Evaluation of faculty performance: Key issues. In Peter Seldon, Changing practices in faculty evaluation (pp. 79-85). San Francisco: Jossey-Bass.

Astin, A. W., & Lee, C. B. T. (1967). Current practices in the evaluation and training of college teaching. In C. B. Lee. Improving College Teaching, Washington, D.C.: American Council on Education.

Avsar, K. M. (1980). An exploratory study of students' written responses to foreign teaching assistants presentation. Dissertation Abstracts International, 41, 11A. (University Microfilms No. 81-09,452)

Bailey, K. M. (1977). The ESL service courses at UCLA: A progress report. In J. Povey (ed.). Working Papers in Teaching English as a Second Language, 11, 1-16.

Bailey, K. M. (1982). Teaching in a second language: The communicative competence of non-native speaking teaching assistants. Dissertation Abstracts International, 43, 10A. (University Microfilms No. 83-06,009)

Bailey, K. M. (1983). Foreign teaching assistants at U.S. universities: Problems in interaction and communication. TESOL Quarterly, 17, 308-310.

Bailey, K. (1984). A typology of teaching assistants. In K. Bailey, F. Pialorski, and J. Zukowski/Faust (Eds.). Foreign Teaching Assistants in U.S. Universities. (pp 110-125) Washington, DC: National Association for Foreign Student Affairs.

Bailey, K. M., Pialorsi, F., & Zukowski-Faust, J. (Eds.) (1984). Foreign Teaching Assistants in U. S. University. National Association for Foreign Student Affairs.

Baldwin, R. G. (1983). "Variety and productivity in faculty careers: Issues in faculty personnel policies." New Directions for Higher Education, No. 41, Fuller. San Francisco: Jossey-Bass.

Banz, M. L., & Rodgers, J. L. (1985). Dimensions underlying student ratings of instruction: A multidimensional scaling analysis. American Educational Research Journal, 22, 267-272.

Barlow, G. A. (1985). Changes in student perceptions of teaching assistants and in self-perceptions of teaching assistants in a college biology course. Dissertation Abstract International, 46, 12A6. (University Microfilms No. 83-06,009)

Barrus, J. L., Armstrong, T. R., Renfren, M. M., & Garrard, V. G., (1974). Preparing teaching assistants. Journal of College Science Teaching, 3, 350-352.

Basow, S. A., & Distenfeld, M. S. (1985). Teacher expressiveness: More important for male teachers than female teachers? Journal of Educational Psychology, 77(1), 45-52.

Basow, S. A., & Howe, K. G. (1979). Sex bias and career evaluations by college women. Perceptual and Motor Skills, 49, 705-706.

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? Journal of Educational Psychology, 79(3), 308-314.

Becker, W. H. (1981). A collegial approach to faculty development. Liberal Education, 67, 19-35.

Bejar, I., & Doyle, K. (1978). Relationship of curriculum area and course format with student rating of instruction. American Educational Research Journal, 15(4), 483-488.

Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. Journal of Educational Psychology, 74(2), 170-179.

Bennett, W. J. (1984). To reclaim a legacy: Text of report on humanities in education. Chronicle of Higher Education, 28, 17-21.

Bennett, J. S. (1985). "Periodic evaluation of tenured faculty performance: In leadership and institutional renewal." New Directions for Higher Education, No. 49. San Francisco: Jossey-Bass.

Bennett, J. B., & Chater, S. S. (1984). Evaluating the performance of tenured faculty members. Educational Record, 65, 38-41.

Bergquist, W. H., & Phillips, S. R. (1975). Components of an effective faculty development program. Journal of Higher Education, 46, 177-211.

Bernhardt, E. (1987). Training foreign teaching assistants: Cultural differences. College Teaching, 35(2), 67-69.

Bevan, J. M. (1985). "Who has the role of building in incentives? Incentives for faculty vitality." New Directions for Higher Education, No. 51, Baldwin. San Francisco: Jossey-Bass.

Blackburn, R. T. (1979). Academic careers: Patterns and possibilities. Current Issues in Higher Education, 2, 25-27.

Blackburn, R. T., Behymer, C. E., & Hall, D. E. (1978). Correlates of faculty publications. Sociology of Education, 51, 132-141.

Blackburn, R. T., & Clark, M. J. (1975). An assessment of faculty performance: Some correlates between administrator, colleague, student, and self-ratings. Sociology of Education, 48, 242-256.

Bland, C., & Schmitz, C. C. (1988). Faculty vitality on review. Journal of Higher Education, 59(2), 191-224.

Blau, P. (1973). The organization of academic work. New York: Wiley.

Bobbitt, K. C. (1985). Systematic faculty evaluation: A growing critical concern. Journal of Nursing Education, 24(2), 86-88.

Boland, D. L., & Sims, S. L. (1988). A comprehensive approach to faculty evaluation. The Journal of Nursing Education, 27(8), 354-358.

Bolden, S. H. (1980). Current status, policy and criteria for the evaluation of tenured faculty in Alabama's four-year public institutions of higher education (Doctoral dissertation, Auburn University, 1979). Dissertation Abstracts International. 40, 10A.

Bos, R., Zakrajesk, D. B., Wolf, V., Stoll, S. (1979). Teaching assistant traits: Their influence on student ratings. Improving College and University Teaching, 28(4), 179-185.

Boyer, E. L. (1987). College: The undergraduate experience in america. New York: Harper and Row.

Brandenburg, D. C., & Aleamoni, L. M. (1976). Illinois course evaluation questionnaire (CEQ): Results interpretation manual. Form 73. Urbana, Ill.: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Brandenburg, D. C., Slinde J. A., & Batista, E. E. (1977). Student ratings of instruction: Validity and normative interpretations. Research in Higher Education, 7, 67-78.

Braskamp, L. A., Brandenburg, D. C., & Ory, J. C. (1984). Evaluating teaching effectiveness: A practical guide. Beverly Hills: Sage.

Braskamp, L., Fowler, D. L., & Ory, J. C. (1984). Faculty development and achievement: A faculty's view. Review of Higher Education, 7, 205-222.

Bray, J. H., & Howard, G. S. (1980). Methodological considerations in the evaluation of a teacher-training program. The American Psychological Association, 72(1), 62-70.

Briton, D., & Gaskill, W. H. (1979). A language skills orientation program for foreign teaching assistants. In J. Povey (Ed.). Workpapers in Teaching English as a Second Language, 13, 49-67.

Brown, D. L., Faculty ratings and student grades: A university-wide multiple regression analysis. Journal of Educational Psychology, 68, 573-578.

Buckenmeyer, J. A. (1972). Preparing graduate assistants to teach. Improving the College and University Teaching, 24, 169-171.

Buhl, L. C. (1982). Empowerment in academic cultures: Whose responsibility is it? In S. C. Inglis & S. Scholl. To Improve the Academy: P.O.D. Resources, Orinda, Calif.: The Professional and Organizational Development Network in Higher Education.

Byrd, P. (1987). Academic sub-cultures within U.S. higher education: Implications for FTA training of differences in teaching styles and methods. Working papers of the First Wyoming Institute on Foreign TA Training (pp. 2-10) University of Wyoming, Laramie, Wyoming.

Byrd, P., Constantinides, J. C., & Pennington, M. C. (1988). The Foreign Teaching Assistant's Manual. McMillan Publishing Company.

Caramagno, T. (1981, April). The TA as model and mirror. The TA at UCLA Newsletter, 7, p. 2.

Carleson, (1985). Research productivity and scholarly accomplishment. College Teaching, 33, 65-67.

Carroll, B. J. (1980). Testing communicative performance: An interim study. Oxford: Pergamon Press.

Centra, J. A. (1972). Strategies for improving college teaching. Washington, D.C.: American Association for Higher Education.

Centra, J. A. (1973a). Self-ratings of college teachers: A comparison with student ratings. Journal of Educational Measurement, 10, 287-295.

Centra, J. A. (1973b). Strategies for improving college teaching. Vol. 8. Washington, D.C. (ERIC Document Reproduction Service No. ED 071 616).

Centra, J. A. (1973c). Effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology, 6(3), 39-41.

Centra, J. A. (1975). Colleagues as raters of classroom instruction. Journal of Higher Education, 46, 327-337.

Centra, J. A. (1976). Development practices in U. S. colleges and university (PR-76-30). Princeton, New Jersey: Educational Testing Service.

Centra, J. A. (1977). How universities evaluate faculty performance: A survey of department heads. (Research Report no. 75-56R). Princeton, N.J.: Graduate Record Examination Program, Educational Testing Service.

Centra, J. A. (1978). Types of faculty development programs. Journal of Higher Education, 49, 11-62.

Centra, J. A. (1979a). Determining faculty effectiveness: Assessing teaching, research, and service for personnel decisions and improvements. San Francisco: Jossey-Bass.

Centra, J. A. (1979b). Research productivity and teaching effectiveness. Research in Higher Education, 18, 379-389.

Centra, J. A. & Creech, F. R. (1976). The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness. (PR-76-1). Princeton, N.J.: Educational Testing Service.

Centra, J. A., Froh, R. C., Gray, P. J., & Lambert, L. M. (1989). A guide to evaluating teaching for promotion and tenure. Syracuse Center for Instructional Development, Syracuse University: Copley Publishing Group Massachusetts.

Chabotar, K. J. (1989). Financial ratio analysis comes to nonprofit. The Journal of Higher Education, 60(2), 188-208.

Chadwick, K. C., & Ward, J. (1987). Determinants of consumer satisfaction with education: Implications for college and university administrators. College and University, 62(3), 236-246.

Chait, R. (1979). "Academic management: What it should be." New Directions for Higher Education, 28, 35-49.

Chait, R., & Ford, A. T. (1982). Beyond traditional tenure. San Francisco: Jossey-Bass.

Choy, C. (1969). The relationship of college teacher effectiveness to conceptual systems orientation and perceptual orientation. Dissertation Abstracts International, 41, 11A. (University Microfilms No. 41-09,291).

Clark, M. J., (1973). Organizational stress and professional performance among faculty members at a small college. Dissertation Abstracts International, 37, 08A. (University Microfilms No. 83-06,009).

Clark, K. W., & Keller, R. J. (1954). Student ratings of college teaching. In R. Eckert & R. J. Keller (Eds.), A university looks at its program, (pp. 197-212). Minneapolis: University of Minnesota.

Cohen, P. A. (1980a). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. Research in Higher Education, 13(4), 321-341.

Cohen, P. A. (1980b). A meta-analysis of the relationship between student rating of instruction and student achievement. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multi-section validity studies. Review of Educational Research, 51(3), 281-309.

Cohen, S. A., & Berger, W. G. (1970). Dimensions of students' ratings of college instructors underlying subsequent achievement on course examinations. Proceedings of the 78th Annual Convention of the American Psychological Association, 5, 605-606.

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

Cohen, P. A., & Mckeachie, W. J. (1960). The role of colleagues in the evaluation of college teaching. Improving College and University Teaching, 28(4), 147-14.

Collette, A. T., & Chiappetta, E. L. (1986). Science instruction in the middle and secondary schools. Columbus, Ohio, Charles E. Merrill Publishing Company.

Constantinides, J. C. (1987). The foreign TA problem: An update. NAFSA Newsletter, 3-6.

Cooper, C. R., & Petrosky, A. (1976). Secondary school students' perceptions of math teachers and math classes. Mathematics Teacher, 69(3), 266-233.

Costin, F., Greenough, W. T., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 41, 511-535.

Creswell, J. W. (1985). How the standards for faculty evaluation have changed. The Review of Education, 21-24.

Crittenden, K. S. (1975). Some remarks on "student ratings": The validity problem. American Educational Research Journal, 12(4), 429-433.

Crooks, T. J., & Kane, M. T. (1981). The generalizability of student ratings of instructors: Item specificity and section effects. Research in Higher Education, 15, 305-313.

Dalgaard, K. A. (1982). Some effects of training on teaching effectiveness of untrained university teaching assistants. Research in Higher Education, 17, 39-50.

Damarin, S. & West, G. (1979, June-July). Preparation of foreign graduate students to teach mathematics: An experimental course. American Mathematical Monthly, 86, 494-497.

Davis, B. K. (1984). A study of the effectiveness of training for foreign teaching assistants. Dissertation Abstracts International, 45, 6A. (University Microfilms No. 84-18,931)

Dege, D. B. (1981, May). Format and evaluation of the cross-cultural component of foreign teaching assistant program. Paper presented at the annual meeting of the International Communication Association, Minneapolis, Minnesota.

Delaney, E. L. (1976). The relationship of student ratings of instruction to student, instructor and course characteristics. Paper read at the annual meeting of the American Educational Research association. (ERIC Document Reproduction Service No. ED 126 140).

Delaney, E. L. (1977). The influence of teaching experience and instructional development activities on student ratings of instruction obtained by beginning professors at a large university. Dissertation Abstracts International, 32, 2A. (University Microfilms No. 64-16,331).

Deming, W. E. (1972). Memorandum on teaching. American Statistician, 26, 47.

Dennis, M. J. (1986). The effects of decreased federal funding on higher education. College and University, 62(1), 48-54.

Diamond, R. M. & Gray, P. (1987). National study of teaching assistants. Syracuse Center for Instructional Development, Syracuse University.

Donato, R. D. (1983). TA Training and supervision: A checklist for an effective program. ADFL bulletin, 15(1), 34-36.

Dowell, D. A., & Neal, J. A. (1982). A selective review of the validity of student ratings of teaching. Journal of Higher Education, 53, 51-62.

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

Doyle, K. O., Jr., & Whitely, S. E. (1974). Student ratings as criteria for effective teaching. American Educational Research Journal, 11, 259-274.

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

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

DuCette, J., & Kenney, J. (1982). Do grading standards affect student evaluation of teaching? Some new evidence on an old question. Journal of Educational Psychology, 74, 308-314.

Eble, K. E. (1970). The recognition and evaluation of teaching. Project to improve college teaching. Salt Lake City, Utah.

Eble, K. E., & McKeachie, W. J. (1985). Improving Undergraduate Education Through Faculty Development: An Analysis of Effective Programs and Practices. San Francisco: Jossey-Bass.

Elmore, P. B., & LaPointe, K. A. (1974). Effects of teacher sex, and student sex on the evaluation of college instructors. Journal of Educational Psychology, 66, 386-389.

Elmore, P. B., & LaPointe, K. A. (1975). Effects of teacher sex, student sex, and teacher warmth on the evaluation of college instructors. Journal of Educational Psychology, 67, 368-374.

Elmore, P. B., & Pohlmann, J. T. (1976). Effect of teacher, student, and class characteristics on the evaluation of college instructors. (Technical Rep. 2.1-76). Carbondale, Ill,: Student Affairs Research and Evaluation Center, Southern Illinois University.

Elmore, P. B., & Pohlmann, J. T. (1978). Effect of teacher, student, and class characteristics on the evaluation of college instructors. Journal of Educational Psychology, 70, 187-192.

Erdle, S., & Murray, H. G. (1986). Inter-faculty differences in classroom teaching behaviors and their relationship to student instructional ratings. Research in Higher Education, 24, 115-127.

Erdle, S., Murray, H. G., & Ruston, J. P. (1985). Personality, classroom behavior and student ratings of college teaching effectiveness: A path analysis. Journal of Educational Psychology, 77(4), 394-407.

Etaugh, C., & Riley, S. (1983). Evaluating competence of women and men: Effects of marital and parental status and occupational sex-typing. Sex Roles, 9(9), 943-952.

Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. Research in Higher Education, 4, 69-111.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. Research in Higher Education, 18, 3-124.

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. Research in Higher Education, 24, 139-213.

Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. Research in Higher Education, 26(3), 227-298.

Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? Research in Higher Education, 28(4), 291-344.

Finkelstein, M. J. (1984). The American Academic Profession: A Synthesis of Social Scientific Inquiry Since World War II. Columbus OH; Ohio State University.

Finn, Jr. C. E. (1984). Trying higher education: An eight count indictment. Change, 28-51.

Foreign teaching assistants. (October 17, 1985). Chronicle of Higher Education, p. 9.

Foreign Teaching Assistants in higher education. (December 20, 1986). Chronicle of Higher Education, p. 23.

Franck, R. M. (1982, Summer). Foreign TAs: A course in communication skills. Improving College and University Teaching, 30, 111-114.

Franck, R. M., & Samaniego, F. A. (1981). The supervision of teaching assistants: A new use of videotape. Modern Language Journal, 65, 273-280.

Fisher, M. (1985). "Rethinking the Foreign TA problem." New Directions for Teaching and Learning, No. 22, 63-73.

Frey, P. W. (1975). Student ratings of teaching: Validity of several rating factors. Science, 182, 83-85.

Frey, P. W., Leonard, D. W., & Beatty, W. W. (1973). Student ratings of instruction: Validation research. American Educational Research Journal, 12(4), 435-447.

Fulton, O., & Trow, M. (1974). Research activity in american higher education. Sociology of Education, 47, 29-73.

Gaff, J. G. (1976). Toward faculty renewal. San Francisco: Jossey-Bass.

Gaff, S. S., C. Festa, & Gaff, J. G. (1978). Professional development: A guide to resources. New Rochelle, N. Y.: Changed Magazine Press.

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

Gage, N. L. (1974). Students' ratings of college teaching: Their justification and proper Use. In N. S. Clasman, & B. R. Killait (eds.), Second UCSB conference on effective teaching. Santa Barbara: University of California.

Galm, J. A. (1985). Welcome to post-tenure review. College Teaching, 33, 65-67.

Garland, T. N. (1983). A training program for graduate teaching assistants the experience of one department. Teaching Sociology, 10, 487-503.

Gill, D. H. (1987). "Instructional evaluation as a feedback process." New Directions for Teaching and Learning, No. 31, San Francisco: Jossey-Bass.

Gillespie, J. K. (1988). Foreign and U.S. teaching assistants: An analysis of verbal and nonverbal classroom interaction. (Doctoral dissertation, University of Illinois, 1987). Dissertation Abstracts International, 49, 9A. (88-23,135)

Gillmore, G. M. (1983-1984). Student ratings as a factor in faculty employment decisions and periodic review. The Journal of College and University Law, 10(4), 56-76.

Gillmore, G. M., & Brandenburg, D. C. (1974). Would the proportion of students taking a class as a requirement affect the student rating of the course? (Research Report No. 347). Urbana, Ill.: Measurement and /Research Division, Office of Instructional Resources, University of Illinois.

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. (1978). The generalizability of student ratings of instruction: Estimation of the teacher and course components. Journal of Educational Measurement, 15(1), 1-13.

Glueck, W. F., & Jauch, L. R. (1975). Sources of research ideas among productive scholars: Implications for administrators. Journal of Higher Education, 46, 103-140.

Goepper, J. B., & Knorre, M. (1980). Pre-and in-service training of graduate teaching assistants. Modern Language Journal, 64, 446-450.

Good, T. L., & Brophy, J. E. (1984). Looking in classrooms (3rd ed. Handbook of Educational Research.). New York: Harper and Row.

Gottschalk, E. C. (1985). Foreign students-teachers faulted for lack of fluency in English. Wall Street Journal.

Grasha, A. F. (1977). Assessing and developing faculty performances: Principles and models. Cincinnati, Ohio: Communication and Education Associates.

Grush, J. E., & Costin, F. (1975). The student as consumer of the teaching process. American Educational Research Journal, 12, 55-66.

Gurnick, D. (1981). And gladly would he learn and gladly teach: The difficult role of the teaching assistant at UCLA. The TA at UCLA Newsletter, 1, p. 3.

Guthrie, E. R., & Costin, F. (1954). The evaluation of teaching: A progress report. Seattle: University of Washington.

Gutie'rrez, J. R. (1987, September). Making better use of the native teaching assistant: Curricular implications. ADFL Bulletin, 19(1), 28-29.

Hammons, J. (1983). "Faculty Development: A necessary corollary to faculty evaluation." In Evaluating Faculty and Staff. New Directions for Community Colleges, No. 41, San Francisco: Jossey-Bass.

Harris, M. B. (1975). Sex role stereotypes and teacher evaluations. Journal of Educational Psychology, 67, 71-756.

Harris, M. B. (1976). The effects of sex, sex-stereotyped descriptions and institution on evaluations of teachers. Sex Roles, 2, 15-21.

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

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

Heller, S. (1984a, September 19). Two years after laying off 58 faculty members, Temple University still feels the after shocks. Chronicle of Higher Education, pp. 23, 25.

Heller, S. (1984b, October 29). Growing pains at Texas Tech: Faculty and president clash over tenure. Chronicle of Higher Education, pp. 12-22.

Heller, S. (1985a, January 16). Review of tenured professors will not work if it's viewed as a threat, administrators say. Chronicle of Higher Education, pp. 29-31.

Heller, S. (1985b, January, 25). Union contracts said to permit faculty review. Chronicle of Higher Education, p. 30.

Heller, S. (1985c, May 5). College try tests and training to make sure foreign TA's can be understood. Chronicle of Higher Education, 13, pp. 32-33.

Hendrickson, R. M., & Lee, B. A. (1983). Academic employment and retrenchment: Judicial review and administrative action. (ASHE -ERIC Higher Education Research Report No.8.) Washington, D.C.: Association for the Study of Higher Education. (ERIC Document Reproduction Service No. ED 240 973).

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

Hildebrand, M., Wilson, R. C., & Dienst, E. R. (1971). Evaluating university teaching. Berkeley, Calif.: Center for Research and Development in Higher Education, University of California at Berkeley.

Hoffman, C. (1975). American Historical Association Faculty Development Program: Planning and Implementation. Paper presented at the annual meeting of the American Historical Association, Atlanta, Georgia.

Holahan, C. K., & Stephen, C. W. (1981). When beauty isn't talent: The influence of physical attractiveness, attitudes toward women, and competence on impression formation. Sex Roles, 7, 867-876.

Holbert, B. J. (1985). Differences in scholarly productivity for faculty employed in comprehensive colleges and universities (Doctoral dissertation, University of Michigan, 1985). Dissertation Abstracts International, 42, 44095A.

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. Journal of Educational Psychology, 77, 187-196.

Howard, G. S., & Maxwell, S. E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? Journal of Educational Psychology, 72, 810-820.

Howard, G. S., & Maxwell, S. E. (1982). Do grades contaminate student evaluations of instruction? Research in Higher Education, 16, 175-188.

Hoyt, D. P. (1974). Interrelationships among instructional effectiveness, publication record, and monetary reward. Research in Higher Education, 2, 81-89.

Hoyt, D. P., & Howard, G. S. (1978). The evaluation of faculty development programs. Research in Higher Education, 8, 25-38.

Huck, S. W., Cormier, W. H., & Rounds, W. G. Jr. (1974). Reading statistics and research. New York: Harper and Row Publishers.

Hudak, M. A., & Anderson, D. E. (1984). Teaching style and student ratings. Teaching of Psychology, 11, 177-178.

Jencks, C., & Riesman, D. (1969). The academic revolution. Garden City, N.Y.: Doubleday Anchor.

Johnson, M., & Kasten, K. (1983). Meritorious work and faculty rewards: An empirical test of the relationship. Research in Higher Education, 19(1), 49-71.

Kaschak, E. (1978). Sex bias in student evaluations of college professors. Psychology of Women Quarterly, 3, 235-243.

Kaschak, E. (1981). Another look at sex bias in students' evaluations of professors: Do winners get the recognition that they have been given? Psychology of Women Quarterly, 5, 767-772.

Kasten, K. L. (1984). Tenure and merit pay as rewards for research, teaching, and service at a research university. Journal of Higher Education, 55(4), 501-514.

Kleinfeld, J. G., McDearmid, G. W., Grubis, S., & Parrett, W. (1984). Doing research on effective cross-cultural teaching: The teacher tale. Peabody Journal of Education, 61, 86-106.

Kounin, J. S. (1970). Discipline and group management in classrooms. New York: Holt, Rinehart & Winston.

Kulik, J. A., & Kulik, C. L. C. (1974). Student ratings of instruction. Teaching of Psychology, 1, 51-57.

Ladd, E. C. (1979). The work experience of American college professors: Some data and an argument. Current Issues in Higher Education, 34, 3-12.

Lawrence, J. H. (1988). Age as a predictor of faculty productivity. Journal of Higher Education, 9(1), 23-38.

Lee, B. A. (1985). Raising the hurdles: Judicial reaction to heightened standards for promotion and tenure. West's Education Law Reporter, 20, 357-364.

Lehman, H. (1968). The creative production rates of present versus past generations of scientists. In B.L. Neugarten (Ed.), Middle age and aging. Chicago: University of Chicago Press, pp. 99-105.

Leventhal, L., Perry, R., & Abrami, P. (1977). Effects of lecturer quality and student perception of lecturer's experience on teacher ratings and student achievement. Journal of Educational Psychology, 69, 360-374.

Levinson-Rose, J., & Menges, R. J. (1981). Improving college teaching: A critical review of research. Review of Educational Research 51, 403-434.

Lewis, W. R., & Orvis, C. C. (1973). A training system for graduate student instructors of introductory economics at the University of Minnesota. Journal of Economic Education, 5, 38-46.

Lindquist, J. (1978). "Social learning and problem-solving strategies for improving academic performance: Evaluating faculty performance and vitality." New Directions for Institutional Research, No. 20, San Francisco: Jossey-Bass.

Linsky, A. S., & Straus, M. A. (1975). Student evaluations, research productivity, and eminence of college faculty. Journal of Higher Education, 46, 89-102.

Lnenicka, W. J. (1972). Are teaching assistants teachers?. Improving College and University Teaching, 20, 97.

Lombardo, J., & Tocci, M. (1979). Attribution of positive and negative characteristics of instructors as a function of attractiveness and sex of instructor and sex of subject. Perceptual and Motor Skills, 48, 491-494.

Long, J. S. (1978). Productivity and academic position in the scientific career. American Sociological Review, 43, 889-908.

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

Lowman, J. (1984). Mastering the techniques of teaching. San Francisco: Jossey-Bass.

Magnusen, K. O. (1987). Faculty evaluation, performance, and pay. Journal of Higher Education, 8(5), 516-529.

Marsh, H. W. (1976). The relationship between background variables and students' evaluations of instructional quality. (OIS 76-9). Los Angeles, CA: Office of Institutional Studies, University of Southern California.

Marsh, H. W. (1977). The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal, 14, 441-447.

Marsh, H. W. (1980a). The influence of student, course, and instructor characteristics in evaluations of university teaching. American Educational Research Journal, 17, 219-237.

Marsh, H. W. (1980b). Research on students' evaluations of teaching effectiveness: A reply to Vecchio. Instructional Evaluation, 4, 5-13.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. Journal of Educational Psychology, 76(5), 707-754.

Marsh, H. W., & Ahocevar, D. (1984). The factorial invariance of student evaluations of college teaching. American Educational Research Journal, 21, 341-366.

Marsh, H. W., & Overall, J. U. (1979). Long-term stability of students' evaluations: A note on Feldman's "Consistency and variability among college students in rating their teachers and courses." Research in Higher Education, 10 139-147.

Marsh, H. W., & Overall, J. U., & Kesler, S. P. (1979). Class size, students' evaluations, and instructional effectiveness. American Educational Research Journal, 16, 57-69.

Marsh, H. W., Overall, J. U. & Thomas (1976). The relationship between students' evaluations of instruction and expected grade. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 126 140).

Marsh, H. W., & Ware,Jr. J. E. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. Journal of Educational Psychology, 74(1), 126-134.

Martinazzi, R. S. (1989). A study of faculty performance evaluation at a comprehensive public university (Doctoral dissertation, University of Pittsburgh, 1987). Dissertation Abstracts International, 50, 3A.

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. Academe, 65, 384-397.

McKeachie, W. J. (1986). A guidebook for the beginning college teacher. Lexington, MA: Heath.

McKeachie, W. J., & Lin, Y. G. (1975). Use of student ratings in evaluation of college teaching. Ann Arbor, MI: National Institute of Education.

McKeachie, W. J., Lin, Y. G., Daugherty, M., Moffett, M., Neigler, C., Nork, J., Walz, M., & Baldwin, R. (1980). Using student ratings and consultation to improve instruction. British Journal of Educational Psychology, 50, 168-174.

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

McKeachie, W. J., Lin, Y. G., Moffett, M., & Daugherty, M. (1978). Effective teaching: Facilitative versus directive style. Teaching of Psychology, 5, 193-194.

Mestenhauser, J. & Barsig, D. (1978). Foreign student advisors as teachers: Learning with foreign students. Washington, D.C., National Association for Foreign Student Affairs.

Miller, M. T. (1971). Instructor attitudes toward, and their use of, student ratings of teachers. Journal of Educational Psychology, 62(3), 235-239.

Millman, J. (1981). Handbook of teacher evaluation. Beverly Hills: Cross-Magan Publishers.

Moffatt, M. (1989, October,4). [Opinion]. The Chronicle of Higher Education, p. 8.

Monrgenstern, R. E. (1969). The relationship between two modes of inter-personal conditions and college teacher effectiveness. Dissertation Abstracts International, 41, 10B. (University Microfilms No. 43-69,291).

Moses, I. (1985). What academics think about regular reviews of performance. Vestes, 28(1), 34-40.

Muhlestein, L. D., & DeFacio, B. (1974). Teaching graduate teaching assistants to teach. American Journal of Physics, 42, 384-386.

Murray, H. G. (1984). The impact of formative and summative evaluation of teaching in North American universities. Assessment and Evaluation in Higher Education, 9(2), 117-132.

Murray, H. G., & Newby, W. G. (1982). Faculty attitudes toward evaluation of teaching at the University of Western Ontario. Assessment and Evaluation in Higher Education, 7(2), 144-151.

Nevill, D. D., Wareses, W. B., & Smith, A. B. (1978). A comparison of student ratings of teaching assistants and faculty members. American Educational Research Journal, 15, 25-37.

Null, E. J., & Nicholson, E. W. (1972). Personal variables of students and their perception of university instructors. College Student Journal, 6, 6-9.

O'Connell, C. (1983). College policies off-target in fostering faculty development. Journal of Higher Education, 54(6), 662-675. Office of the Vice President for Personnel Services, Oregon State University, October 1989.

Office of Academic Affairs Oregon State University (1989, September). Preparing The Future: Revised Strategic Plan.

Olswang, S. G. (1982-1983). Planning the unthinkable: Issues in institutional reorganization and faculty reductions. Journal of College and University Law, 9, 431-449.

Olswang, S. G., & Fantel, J. I. (1980-1981). Tenure and periodic performance review: Compatible legal and administrative principles. Journal of College and University Law, 7, 1-30.

Olswang, S. G. & Lee, B. A. (1984). Faculty freedoms and institutional accountability: Interactions and conflicts. (ASHE-ERIC Higher Education Research Report No. 5.) Washington, D.C.: Association for the Study of Higher Education. (ERIC Document Reproduction Service No. ED 252 170).

Ormrod, J. E. (1986). Predictors of faculty dissatisfaction with an annual performance evaluation. The Journal of the College and University Personnel Association, 37, 13-18.

Overall, J. U., & Marsh, H. W. (1979). Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology, 71, 856-865.

Paludi, M. A., & Bauer, W. D. (1983). Goldberg revisited: What's in an author's name. Sex Roles, 9, 387-390.

Pambookian, H. S. (1976). Discrepancy between instructor and student evaluation of instruction: Effect on instructor. Instructional Science, 5, 63-7.

Paulston, C. B. (1974). Linguistic and communicative competence. TESOL Quarterly, 8(4), 347-362.

Pellino, R. T., Blackburn, R. T., & Boberg, A. L. (1984). The dimensions of academic scholarship: Faculty and administrator views. Research in Higher Education, 20(1), 103-115.

Peterson, H. L. (1981). Coping with the 1980s: An organizational response to retrenchment. (ERIC Document Reproduction Service no. ED 203 829).

Piper, D. (1983, July). An Interview of Department Heads in Institutions Across the United Kingdom. Paper presented at the Ninth International Conference on University Teaching, Dublin Ireland.

Pohlmann, J. T. (1975). A multivariate analysis of selected class characteristics and student ratings of instruction. Behavioral Research, 10(1), 81-91.

Rava, S. (1987). Training teaching assistants. ADFL Bulletin, 19(1), 26-27.

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

Remmers, H. H. (1949). To what extent do grades influence student ratings of instructors? Journal of Educational Research, 21, 314-316.

Remmers, H. H. (1963). Teaching methods in research on teaching. In N. L. Gage (Ed.), Handbook on teaching. Chicago: Rand McNally.

Renfrew, M. M. (1978). The training of teaching assistants in chemistry: A survey. Journal of Chemical Education, 55, 386-388.

Rice, D. S. (1979, May). A description of a model program for orienting the new foreign teaching assistant. Paper read at thirty first annual conference of the National Association for Foreign Student Affairs, Phoenix, Arizona.

Rice, R. E., & Davis, M. L. (1979). Program coordination of academic planning and professional development (Final report to Kellogg Foundation). Grand Rapids, Michigan. (ERIC Document Reproduction Service No. ED. 172 688).

Rotem, A., & Glasman, N. S. (1979). On the effectiveness of students' evaluative feedback to university instructors. Review of Educational Research, 49, 497-511.

Rounds, P. R. (1985). Talking the mathematics through: Disciplinary transaction and socio-educational interaction. Dissertation Abstracts International, 46, 3338A. (University Microfilms No. 86-00, 543).

Rounds, P. R. (1987). Characterizing successful classroom discourse for NNTAs teaching assistant training. TESOL Quarterly, 26(4), 643 -671.

Royal, S. J. (1972). Train them first. Improving College and University Teaching, 17, 44-47.

Russo, G. (1982). Training for TAs. Improving College and University Teaching, 31, 171-174.

Schank, R. (1984). The cognitive computer: Reading. Boston: Addison-Wesley.

Scheurich, Graham, Drolette (1983). Expected grades versus specific evaluations of the teacher as predictors of students' overall evaluation of the teacher. Research in Higher Education, 19(2), 159-173.

Schwab, D. P. (1975). Course and student characteristic correlates of the course evaluation instrument. Journal of Applied Psychology, 60, 742-747.

Seldin P. (1977). Faculty growth contracts. London, England: University of London, Institute of Education. (ERIC Document Reproduction Service No. ED 157 469).

Seldin, P. (1980). Successful faculty evaluation programs. New York: Coventry Press.

Seldin, P. (1981). Faculty development programs. Exchange, 6, 17-21.

Seldin, P. (1982). Self-assessment of college teaching. Improving College and University Teaching, 30(2), 70-74.

Seldin, P. L. (1984). Changing practices in faculty evaluation. San Francisco: Jossey-Bass.

Shahenayati, Z. S. (1987). A comparison of native and non-native English-speaking teaching assistants. Dissertation Abstracts International, 49, 5A. (University Microfilms NO. 88-12,410).

Sheehan, D. S. (1975). On the invalidity of student ratings for administrative personnel decisions. Journal of Higher Education, 6, 687-700.

Shirvani (1987). A comparison of native and non-native english-speaking teaching assistants. Disser-tation Abstracts International, 49, 7A. (University Microfilms NO. 8713-979).

Siebring, B. R. (1972). A training program for teaching assistants. Improving College and University Teaching, 20, 98-99.

Singhal, S. (1968). Illinois course evaluation questionnaire items by rank of instructor, sex of the instructor, and sex of the student. (Research Report No. 282). Urbana-Champaign, Ill.: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Slaughter, S. (1988). Academic freedom and the state: Reflections on the uses of knowledge. Journal of Higher Education, 59(3), 241-262.

Slaughter, S., & Silva, E. T. (1985). Towards a political economy of retrenchment: The American public research universities. Review of Higher Education, 8(4), 295-318.

Smock, H. R., & Crooks, T. J. (1973). A plan for the comprehensive evaluation of college teaching. Journal of Higher Education, 44, 572-586.

Stallings, W. M., & Singal, S. (1970). Some observations on the relationships between research productivity and student evaluations of courses and teaching. The American Sociologist, 5, 143-143.

Stallings, W. M., & Spencer, R. E. (1967). Ratings of Instructors in Accountancy 101 from Videotape Clips. (Research Report No. 265). Urbana: Office of Instructional Resources, Measurement and Research Division, University of Illinois.

Staton-Spicer, A. Q., & Nyquist, J. L. (1979, July). Improving the teaching effectiveness of graduate teaching assistants. Communication Education, 28, 199-205.

Stevens, J. J. (1987). "Using student ratings to improve instruction." New Directions for Teaching and Learning, No. 31, San Francisco: Jossey-Bass.

Stevens, J. J., & Aleamoni, L. M. (1985). The use of evaluative feedback for instructional improvement: A longitudinal perspective. Instructional Science, 13, 285-304.

Stockdale, D. L., & Wochok, Z. S. (1974). Training TA's to teach. Journal of College Science Teaching, 3(5), 345-349.

Stumpf, S. A., & Freedman, R. D. (1979). Expected grade covariation with student ratings of instruction: Individual versus class effects. Journal of Educational Psychology, 71, 293-302.

Stumpf, S. A., Freedman, R. D., & Aguanno (1979). A path analysis of factors often found to be related to student ratings of teaching effectiveness. Research in Higher Education, 11(2), 111-123.

Swanbeck, H. (1981). Foreign TAs experience communication gap in classroom. UCLA Daily Bruin, 26, pp. 9-10.

The National Study of Teaching Assistants (May 18, 1987), The Wall Street Journal, p. 32.

Thomas, D. M. (1980). Performance evaluation: Basis for job motivation. National Association of Secondary School Principals Bulletin, 64(434), 8-11.

Tollefson, N. (1980). The influence of selected student and course characteristics on teacher effectiveness ratings (ERIC Document Reproduction Service No. ED 195 563).

Van Horn, C. (1968). An analysis of the 1968 course and instructor evaluation report. (Institutional Research Bulletin No. 2-68). West Lafayette, IN: Measurement and Research Center, Purdue University.

Voeks, V. W. (1962). Publications and teaching effectiveness: A search for some relationship. Journal of Higher Education, 212-218.

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 International, 29(9-B), 3447.

Ware, J. E., & Williams, R. G. (1975). The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction. Journal of Medical Education, 50, 149-150.

Weerts, R. R., & Whitney, D. R. (1975). Student perceptions of teaching (SPOT): V. Relationships between averaged student responses and selected course characteristics. (Rep. No. 78.) Iowa City, Iowa: Evaluation and Examination Service, University of Iowa.

White, W. F., Hsu, Y. M., & Means, R. S. (1978). Prediction of student ratings of college instructors from multiple achievement test variables. Educational and Psychological Measurement, 38(4), 1077-1083.

Whitman, N., & Schwenk, T. (1982). Faculty evaluation as a means of faculty development. Journal of Family Practice, 14, 1097-1101.

Whitman, N., & Weiss, E. (1982). Faculty Evaluation: The Use of Explicit Criteria for Promotion, Retention, and Tenure. (AAHE -ERIC Higher Education Research Report No. 2). Washington, D.C.: American Association for Higher Education. (ERIC Document Reproduction Service No. ED 221 148).

Wilson, R. C. (1986). Improving faculty teaching: Effective use of student evaluations and consultants. Journal of Higher Education, 57, 196-211.

Wittrock, M.C. (1974a). Learning as a generative process. Educational Psychologist, 11, 87-95.

Wittrock, M.C. (1974b). A generative model of mathematics learning. Journal of Research in Mathematics Education, 5, 181-197.

Wotruba, T. R., & Wright, P. L. (1974). How to develop a teacher-rating instrument: A research approach. Journal of Higher Education, 46(6), 653-663.

Wulff, D. H., Nyquist, J. D., & Abbott, R. D. (1987). "Students' perceptions of large classes." New Directions for Teaching and Learning, No. 31, 17-30.

Zakrajsek D. B., Wolf, V., & Stoll, S. (1985). Teaching assistant traits: Their influence on student ratings. Improving College and University Teaching, 28(4), 179-185.

APPENDICES

Appendix A

The Student Assessment of Teaching Instrument (SATI)

OREGON STATE UNIVERSITY
STUDENT ASSESSMENT OF TEACHING

USE NO. 2 PENCIL ONLY

INSTRUCTOR'S NAME	DEPARTMENT	COURSE NUMBER/TITLE	SECTION #	DATE

THIS QUESTIONNAIRE GIVES YOU AN OPPORTUNITY TO EXPRESS YOUR VIEWS OF THIS COURSE AND THE WAY IT HAS BEEN TAUGHT.

SECTION I: Information for Evaluating Teaching and for Improving Instruction. (Items 1-12)

(PLEASE FILL-IN THE APPROPRIATE RESPONSE, MARK ONLY ONE CIRCLE PER QUESTION)

		NO BASIS FOR OPINION	STRONGLY DISAGREE				STRONGLY AGREE
1	Course objectives and requirements were clearly presented to me.	⓪	①	②	③	④	⑤
2	The Instructor was well prepared and organized.	⓪	①	②	③	④	⑤
3	The Instructor explained the material clearly.	⓪	①	②	③	④	⑤
4	The Instructor was sensitive to my/the class' ability to understand the material.	⓪	①	②	③	④	⑤
5	The Instructor stimulated enthusiasm for the subject matter of the course.	⓪	①	②	③	④	⑤
6	The Instructor provided scheduled office hours or was readily available for consultation with me.	⓪	①	②	③	④	⑤
7	The Instructor was fair and impartial in dealing with me.	⓪	①	②	③	④	⑤
8	The Instructor encouraged me to think for myself.	⓪	①	②	③	④	⑤
9	The examinations were relevant to the reading assignments and to the material presented in class.	⓪	①	②	③	④	⑤
10	The Instructor used good communication skills.	⓪	①	②	③	④	⑤
11	As a result of having this Instructor, I have learned a significant number of new ideas and/or skills.	⓪	①	②	③	④	⑤
12.	All things considered, I was favorably impressed by this instructor.	⓪	①	②	③	④	⑤

SECTION II: DEMOGRAPHICS (ITEMS 13-19)

PLEASE FILL-IN ONE RESPONSE CIRCLE FOR EACH OF THE FOLLOWING ITEMS WHICH BEST DESCRIBES YOUR SITUATION.

13. THE REASON YOU ARE ENROLLED IN THIS COURSE:
    - Ⓐ IT IS REQUIRED.
    - Ⓑ IT IS AN ELECTIVE.

14. GRADE YOU EXPECT TO RECEIVE IN THIS COURSE:
    - Ⓐ A
    - Ⓑ B
    - Ⓒ C
    - Ⓓ D
    - Ⓔ F
    - Ⓕ SAT/PASS
    - Ⓖ UNSAT/NO PASS
    - Ⓗ AUDIT
    - Ⓘ OTHER

15. CLASS STATUS:
    - Ⓐ FRESHMAN
    - Ⓑ SOPHOMORE
    - Ⓒ JUNIOR
    - Ⓓ SENIOR
    - Ⓔ GRADUATE STUDENT
    - Ⓕ OTHER

16. IS THIS COURSE IN YOUR MAJOR?
    - Ⓨ YES
    - Ⓝ NO

17. PERCENT OF THIS CLASS YOU ATTENDED:
    - Ⓐ 0-20%
    - Ⓑ 21-40%
    - Ⓒ 41-60%
    - Ⓓ 61-80%
    - Ⓔ 81-100%

18. OVERALL GRADE POINT AVERAGE:
    - Ⓐ 0-1.49
    - Ⓑ 1.50-1.99
    - Ⓒ 2.00-2.49
    - Ⓓ 2.50-2.99
    - Ⓔ 3.00-3.49
    - Ⓕ 3.50-4.00
    - Ⓖ 1ST QUARTER FRESHMAN

19. SEX:
    - Ⓜ MALE
    - Ⓕ FEMALE

Appendix B

Professional Journals of Education at O.S.U. Kerr Library

* Refereed journals meeting TALAGLIS criteria. The titles, call numbers, and Serials Checkin Service international journal code numbers have been included for each journal for quick reference. This compilation comes from the Kerr Library Support Services Administration who periodically update their list of holdings for library use and to facilitate university wide research.

AS OF 02/20/91 SERIALS CHECKIN SERVICE: SERIALS HOLDINGS LIST SORT: FISL

JOBID=S140 PAGE: 1440
REQUESTOR 02 MV
SHIPTO#: 62332

* FISL: P.EDUC
TITLE NAME: ACTION IN TEACHER EDUCATION
CALL NO : LB_1715.A39

FISL: P.EDUC
TITLE NAME: ADMINISTRATIVE INFORMATION REPORT
CALL NO : LB_2805.A35

* FISL: P.EDUC
TITLE NAME: ADULT EDUCATION QUARTERLY
CALL NO : LC_5201.A26

* FISL: P.EDUC
TITLE NAME: ADULT LEARNING (WASHINGTON, D.C.)
CALL NO : LC_5251.A481

FISL: P.EDUC
TITLE NAME: ADULT LITERACY AND BASIC EDUCATION
CALL NO : LC_5201.A45

* FISL: P.EDUC
TITLE NAME: AEDS JOURNAL / ASSOCIATION FOR EDUCATIONAL DATA SYSTEMS.
CALL NO : LB_2846.A78

* FISL: P.EDUC
TITLE NAME: AMERICAN ANNALS OF THE DEAF
CALL NO : HV_2510.A5

* FISL: P.EDUC
TITLE NAME: AMERICAN BIOLOGY TEACHER
CALL NO : QH__301.A5

* FISL: P.EDUC
TITLE NAME: AMERICAN EDUCATIONAL RESEARCH JOURNAL
CALL NO : LB_1028.A5

* FISL: P.EDUC
TITLE NAME: AMERICAN EDUCATOR
CALL NO : L____11.A46

* FISL: P.EDUC
TITLE NAME: AMERICAN EDUCATOR. (BINDING COPY)
CALL NO : L____11.A46

* FISL: P.EDUC
TITLE NAME: AMERICAN JOURNAL OF EDUCATION.
CALL NO : L____11.S37

* FISL: P.EDUC
TITLE NAME: AMERICAN JOURNAL OF MENTAL RETARDATION : AJMR.
CALL NO : LC_4001.A5

* FISL: P.EDUC
TITLE NAME: AMERICAN SCHOOL AND UNIVERSITY
CALL NO : L____11.A49

* FISL: P.EDUC
TITLE NAME: AMERICAN SCHOOL BOARD JOURNAL.
CALL NO : L____11.A5

FISL: P.EDUC
TITLE NAME: ARTS AND ACTIVITIES
CALL NO : L____11.A7

FISL: P.EDUC
TITLE NAME: ASSOCIATION FOR CHILDHOOD EDUCATION INTERNATIONAL. ACEI EXCHANGE.
CALL NO : DISCARD

FISL: P.EDUC
TITLE NAME: ASSOCIATION FOR EDUCATIONAL COMMUNICATIONS & TECHNOLOGY. COMPREHENSIVE MEMBERSHIP
CALL NO : SEE COMMENTS

FISL: P.EDUC
TITLE NAME: ASSOCIATION OF TEACHER EDUCATORS. LIBRARY/INSTITUTIONAL SUBSCRIPTION
CALL NO : SEE COMMENTS

FISL: P.EDUC
TITLE NAME: ATE NEWS LETTER.
CALL NO : LB_1715.A881

* FISL: P.EDUC
TITLE NAME: BEHAVIORAL DISORDERS.
CALL NO : RJ__506.B44 B41

* FISL: P.EDUC
TITLE NAME: BRITISH JOURNAL OF EDUCATIONAL PSYCHOLOGY
CALL NO : LB_1051.A1 B7

* FISL: P.EDUC
TITLE NAME: BRITISH JOURNAL OF EDUCATIONAL STUDIES
CALL NO : L____16.B7

* FISL: P.EDUC
TITLE NAME: CANADIAN JOURNAL OF EDUCATION. REVUE CANADIENNE DE L'EDUCATION.
CALL NO : LA__410.C36

* FISL: P.EDUC
TITLE NAME: CHANGE
CALL NO : LB_2300.C34

FISL: P.EDUC
TITLE NAME: CHANGE. (BINDING COPY)
CALL NO : LB_2300.C34

FISL: P.EDUC
TITLE NAME: CHILD DEVELOPMENT ABSTRACTS AND BIBLIOGRAPHY.
CALL NO : LB_1103.N2

FISL: P.EDUC
TITLE NAME: CHILD DEVELOPMENT. COP.1
CALL NO : LB_1101.C4

* FISL: P.EDUC
TITLE NAME: CHILD STUDY JOURNAL
CALL NO : LB_1101.C43

* FISL: P.EDUC
TITLE NAME: CHILDHOOD EDUCATION. COP.1
CALL NO : LB_1141.C5

FISL: P.EDUC
TITLE NAME: CHILDHOOD EDUCATION. MEMBERSHIP/SUBSCRIPTION
CALL NO : SEE COMMENTS

FISL: P.EDUC
TITLE NAME: CLEARING HOUSE.
CALL NO : L____11.C4

* FISL: P.EDUC
TITLE NAME: COLLEGE AND UNIVERSITY
CALL NO : LB_2301.A49

* FISL: P.EDUC
TITLE NAME: COLLEGE BOARD REVIEW
CALL NO : LB_2353.C48

* FISL: P.EDUC
TITLE NAME: COMMUNITY AND JUNIOR COLLEGE JOURNAL
CALL NO : LB_2300.J8

* FISL: P.EDUC
TITLE NAME: COMMUNITY COLLEGE REVIEW
CALL NO : LB_2328.C57

* FISL: P.EDUC
TITLE NAME: COMMUNITY, TECHNICAL, AND JUNIOR COLLEGE JOURNAL
CALL NO : LB_2300.J8

* FISL: P.EDUC
TITLE NAME: COMPARATIVE EDUCATION REVIEW.
CALL NO : L____11.C57

* FISL: P.EDUC
TITLE NAME: COMPUTERS & EDUCATION
CALL NO : LB_2846.C578

* FISL: P.EDUC
TITLE NAME: CONTINUING HIGHER EDUCATION REVIEW
CALL NO : L____13.N48

* FISL: P.EDUC
TITLE NAME: CONVERGENCE.
CALL NO : LC_5201.C6

FISL: P.EDUC
TITLE NAME: CONVERGENCE. (BINDING COPY)
CALL NO : LC_5201.C6

FISL: P.EDUC
TITLE NAME: COUNSELING AND VALUES
CALL NO : LC__461.N438

* FISL: P.EDUC
TITLE NAME: CUPA JOURNAL (WASHINGTON, D.C. : 1987).
CALL NO : LB_2342.7 J681

FISL: P.EDUC
TITLE NAME: DEAF AMERICAN
CALL NO : HV_2350.D41

FISL: P.EDUC
TITLE NAME: DROPOUT PREVENTION : A BOOK OF SOURCES.
CALL NO : LC__143.D71

* FISL: P.EDUC
TITLE NAME: EDUCATION
CALL NO : L____11.E24

* FISL: P.EDUC
TITLE NAME: EDUCATION AND TRAINING IN MENTAL RETARDATION
CALL NO : HV__894.E4

FISL: P.EDUC
TITLE NAME: EDUCATION & TREATMENT OF CHILDREN.
CALL NO : LB_1051.A2 E31

FISL: P.EDUC
TITLE NAME: EDUCATION AND URBAN SOCIETY
CALL NO : LC_5101.E38

* FISL: P.EDUC
TITLE NAME: EDUCATION DIGEST. COP.1
CALL NO : L____11.E282

* FISL: P.EDUC
TITLE NAME: EDUCATION WEEK
CALL NO : L____11.E2991

* FISL: P.EDUC
TITLE NAME: EDUCATIONAL ADMINISTRATION QUARTERLY.
CALL NO : L____11.E305

* FISL: P.EDUC
TITLE NAME: EDUCATIONAL FORUM
CALL NO : L____11.E31

* FISL: P.EDUC
TITLE NAME: EDUCATIONAL LEADERSHIP
CALL NO : L____11.E324

* FISL: P.EDUC
TITLE NAME: EDUCATIONAL RESEARCH
CALL NO : LB_1028.E3

* FISL: P.EDUC
TITLE NAME: EDUCATIONAL RESEARCH QUARTERLY.
CALL NO : L____11.E345

* FISL: P.EDUC
TITLE NAME: EDUCATIONAL TECHNOLOGY
CALL NO : L____11.E36

FISL: P.EDUC
TITLE NAME: EDUCATIONAL TECHNOLOGY RESEARCH AND DEVELOPMENT : ETR & D.
CALL NO : LB_1043.A781

* FISL: P.EDUC
TITLE NAME: ELEMENTARY SCHOOL JOURNAL
CALL NO : L____11.E6

* FISL: P.EDUC
TITLE NAME: ENGLISH EDUCATION
CALL NO : LA__632.E6

FISL: P.EDUC
TITLE NAME: EPIEGRAM.
CALL NO : ORDERED 08/17/88

FISL: P.EDUC
TITLE NAME: EQUITY & EXCELLENCE.
CALL NO : LB_3062.I5

FISL: P.EDUC
TITLE NAME: EXCEPTIONAL CHILD EDUCATION RESOURCES
CALL NO : LC_3950.E9

FISL: P.EDUC
TITLE NAME: EXCEPTIONAL CHILDREN
CALL NO : LC_3951.E9

FISL: P.EDUC
TITLE NAME: FOCUS ON EXCEPTIONAL CHILDREN
CALL NO : LC_3950.F631

FISL: P.EDUC
TITLE NAME: GIFTED CHILD QUARTERLY
CALL NO : LC_3991.G5

FISL: P.EDUC
TITLE NAME: GIFTED CHILD TODAY
CALL NO : LC_3991.G2

FISL: P.EDUC
TITLE NAME: HARVARD EDUCATIONAL REVIEW.
CALL NO : L____11.H2

FISL: P.EDUC
TITLE NAME: HEALTH EDUCATION
CALL NO : LB_3401.S35

FISL: P.EDUC
TITLE NAME: HIGH SCHOOL JOURNAL
CALL NO : L____11.H48

FISL: P.EDUC
TITLE NAME: HIGHER EDUCATION QUARTERLY
CALL NO : L____16.U55

FISL: P.EDUC
TITLE NAME: HISTORY OF EDUCATION QUARTERLY
CALL NO : L____11.H67

FISL: P.EDUC
TITLE NAME: INDEPENDENT SCHOOL.
CALL NO : LC___47.I5

* FISL: P.EDUC
TITLE NAME: INDUSTRIAL EDUCATION
CALL NO : T___61.I5

FISL: P.EDUC
TITLE NAME: INSTRUCTOR.
CALL NO : L____11.I61

FISL P.EDUC
TITLE NAME: INTERNATIONAL READING ASSOC. BASIC MEMBERSHIP W/4 JOURNALS & W/BOOK CLUB
CALL NO SEE COMMENTS

* FISL: P.EDUC
TITLE NAME: INTERNATIONAL REVIEW OF EDUCATION. INTERNATIONALE ZEITSCHRIFT FUR ERZIEHUNGSWISSENSCHAFT. REVUE INTE
CALL NO : L____10.I87

FISL: P.EDUC
TITLE NAME: INTERVENTION IN SCHOOL AND CLINIC
CALL NO : LC_4001.A3

FISL: P.EDUC
TITLE NAME: JEWISH EDUCATION
CALL NO : LC__701.J4

FISL: P.EDUC
TITLE NAME: JOHN DEWEY SOCIETY. INSTITUTIONAL MEMBERSHIP
CALL NO : SEE COMMENTS

* FISL: P.EDUC
TITLE NAME: JOURNAL OF AMERICAN INDIAN EDUCATION
CALL NO : E____97.J6

* FISL: P.EDUC
TITLE NAME: JOURNAL OF COMPUTER - BASED INSTRUCTION
CALL NO : LB_1028.5 J613

* FISL: P.EDUC
TITLE NAME: JOURNAL OF COOPERATIVE EDUCATION.
CALL NO : LB_1029.C6 J62

FISL: P EDUC
TITLE NAME: JOURNAL OF COOPERATIVE EDUCATION. WITH COOPERATIVE EDUCATION MEMBERSHIP DIRECTORY
CALL NO . SEE COMMENTS

* FISL P EDUC
TITLE NAME: JOURNAL OF EDUCATION. COP.1
CALL NO : L____11.J55

* FISL: P EDUC
TITLE NAME: JOURNAL OF EDUCATIONAL MEASUREMENT.
CALL NO : LB_3051.J6

* FISL: P EDUC
TITLE NAME: JOURNAL OF EDUCATIONAL PSYCHOLOGY.
CALL NO : L____11.J6

* FISL: P EDUC
TITLE NAME: JOURNAL OF EDUCATIONAL RESEARCH
CALL NO : L____11.J63

* FISL: P EDUC
TITLE NAME: JOURNAL OF EXPERIMENTAL EDUCATION
CALL NO : L____11.J75

* FISL: P EDUC
TITLE NAME: JOURNAL OF GENERAL EDUCATION. COP.1
CALL NO : L____11.J64

* FISL: P EDUC
TITLE NAME: JOURNAL OF GENETIC PSYCHOLOGY
CALL NO : L____11.P4

* FISL: P EDUC
TITLE NAME: JOURNAL OF HIGHER EDUCATION
CALL NO : LB_2300.J6

. FISL: P EDUC
TITLE NAME: JOURNAL OF LEARNING DISABILITIES.
CALL NO LB_1134.J6

FISL: P EDUC
TITLE NAME: JOURNAL OF MEMORY AND LANGUAGE
CALL NO : BF_455.A1 J6

FISL: P EDUC
TITLE NAME: JOURNAL OF NEGRO EDUCATION
CALL NO : LC_2701.J6

FISL: P EDUC
TITLE NAME: JOURNAL OF NUTRITION EDUCATION. COP.1
CALL NO : LB_1587.N8 J7

FISL: P EDUC
TITLE NAME: JOURNAL OF PHILOSOPHY OF EDUCATION
CALL NO : L____18.P51

FISL: P EDUC
TITLE NAME: JOURNAL OF READING.
CALL NO : LB_1050.J6

FISL: P EDUC
TITLE NAME: JOURNAL OF READING BEHAVIOR.
CALL NO : LB_1050.J62

FISL: P EDUC
TITLE NAME: JOURNAL OF RESEARCH AND DEVELOPMENT IN EDUCATION
CALL NO : LB_1028.J64

FISL: P EDUC
TITLE NAME: JOURNAL OF RESEARCH ON COMPUTING IN EDUCATION
CALL NO : LB_2846.A78

FISL: P EDUC
TITLE NAME: JOURNAL OF SCHOOL HEALTH
CALL NO : LB_3401.J6

FISL: P EDUC
TITLE NAME: JOURNAL OF SCHOOL PSYCHOLOGY
CALL NO : LB_3013.6 J6

FISL: P.EDUC
TITLE NAME: JOURNAL OF SPECIAL EDUCATION
CALL NO : LC_3951.J6

FISL: P.EDUC
TITLE NAME: JOURNAL OF TEACHER EDUCATION
CALL NO : LB_1705.J6

FISL: P.EDUC
TITLE NAME: LANGUAGE ARTS
CALL NO : L____11.E5

FISL: P.EDUC
TITLE NAME: LANGUAGE, SPEECH AND HEARING SERVICES IN SCHOOLS
CALL NO : LB_1139.L3 L183

FISL: P.EDUC
TITLE NAME: LEARNING.
CALL NO : L____11.L4

FISL: P.EDUC
TITLE NAME: LEARNING. (BINDING COPY)
CALL NO : L____11.L4

FISL: P.EDUC
TITLE NAME: LEARNING DISABILITY QUARTERLY
CALL NO : LC_4704.L424

FISL: P.EDUC
TITLE NAME: LIBERAL EDUCATION
CALL NO : LB_2301.A56

FISL: P.EDUC
TITLE NAME: MEDIA & METHODS
CALL NO : LB_1043.M4

FISL: P.EDUC
TITLE NAME: MIDDLE SCHOOL JOURNAL
CALL NO : L____11.M65

* FISL: P.EDUC
TITLE NAME: NATIONAL ASSOCIATION FOR WOMEN DEANS, ADMINISTRATORS AND COUNSELORS. JOURNAL
CALL NO : LC_1551.N22

FISL: P.EDUC
TITLE NAME: NATIONAL ASSOCIATION OF SECONDARY SCHOOL PRINCIPALS. SUBSCRIPTION
CALL NO : SEE COMMENTS

* FISL: P.EDUC
TITLE NAME: NATIONAL ASSOCIATION OF SECONDARY SCHOOL PRINCIPALS. BULLETIN.
CALL NO : LB_2804.N3

FISL: P.EDUC
TITLE NAME: NATIONAL ASSOCIATION OF SECONDARY SCHOOL PRINCIPALS. CURRICULUM REPORT.
CALL NO : LB_2804.N32

* FISL: P.EDUC
TITLE NAME: NATIONAL ASSOCIATION OF STUDENT PERSONNEL ADMINISTRATORS. NASPA JOURNAL.
CALL NO : LB_2343.N2

FISL: P.EDUC
TITLE NAME: NEA TODAY.
CALL NO : L____13.N4141

* FISL: P.EDUC
TITLE NAME: NEGRO EDUCATIONAL REVIEW
CALL NO : L____11.N3

* FISL: P.EDUC
TITLE NAME: NORTH CENTRAL ASSOCIATION QUARTERLY
CALL NO : L____11.N6

FISL: P.EDUC
TITLE NAME: OREGON EDUCATION
CALL NO : L____13.075

* FISL: P.EDUC
TITLE NAME: PEABODY JOURNAL OF EDUCATION
CALL NO : L____11.P35

FISL: P.EDUC
TITLE NAME: THE PRACTITIONER.
CALL NO : LB_2805.P7

FISL: P.EDUC
TITLE NAME: PSYCHOLOGY IN THE SCHOOLS
CALL NO : LB_1101.P7

FISL: P.EDUC
TITLE NAME: RE:VIEW (ASSOCIATION FOR EDUCATION AND REHABILITATION OF THE BLIND AND VISUALLY IMPAIRED (U.S.)).
CALL NO : HV_1571.E4

FISL: P.EDUC
TITLE NAME: READING RESEARCH AND INSTRUCTION
CALL NO : LB_1050.J65

* FISL: P.EDUC
TITLE NAME: READING RESEARCH QUARTERLY.
CALL NO : LB_1050.R43

* FISL: P.EDUC
TITLE NAME: READING TEACHER.
CALL NO : LB_1573.R4

* FISL: P.EDUC
TITLE NAME: REVIEW OF EDUCATIONAL RESEARCH
CALL NO : L____11.R4

FISL: P.EDUC
TITLE NAME: SCHOLASTIC UPDATE.
CALL NO : AP____2.S37

FISL: P.EDUC
TITLE NAME: SCHOOL AND COMMUNITY
CALL NO : L____11.S29

* FISL: P.EDUC
TITLE NAME: SCHOOL COUNSELOR
CALL NO : LB_1027.5 S28

FISL: P EDUC
TITLE NAME: SCHOOL FOOD SERVICE JOURNAL
CALL NO : LB_3475.A1 S3

FISL: P EDUC
TITLE NAME: SCHOOL SHOP, TECH DIRECTIONS.
CALL NO : T____61.S3

FISL: P EDUC
TITLE NAME: SCHOOL SHOP, TECH DIRECTIONS (MICROFICHE)
CALL NO : T____61.S3

FISL: P EDUC
TITLE NAME: SCHOOLS IN THE MIDDLE / NATIONAL ASSOCIATION OF SECONDARY SCHOOL PRINCIPALS.
CALL NO : LB_1623.S361

FISL: P EDUC
TITLE NAME: SCHOOLTECHNEWS. -- NASSP ED.
CALL NO : LB_1028.3 S361

FISL: P EDUC
TITLE NAME: SOCIOLOGY OF EDUCATION
CALL NO : L____11.J7

FISL: P EDUC
TITLE NAME: TEACHERS COLLEGE RECORD.
CALL NO : L____11.T4

FISL: P EDUC
TITLE NAME: TEACHING AND TEACHER EDUCATION
CALL NO : LB_1025.2 T4154

FISL: P EDUC
TITLE NAME: TEACHING EXCEPTIONAL CHILDREN. COP.1
CALL NO : LC_3950.T4

FISL: P EDUC
TITLE NAME: TEACHING PRE K - 8
CALL NO : LB_1501.E27

FISL: P.EDUC
TITLE NAME: TEACHING SOCIOLOGY
CALL NO : HM____1.T43

FISL: P.EDUC
TITLE NAME: TECHNOLOGY & LEARNING
CALL NO : LB_1028.5 C531

FISL: P.EDUC
TITLE NAME: TECHNOLOGY TEACHER
CALL NO : TT__161.I49

FISL: P.EDUC
TITLE NAME: TECHTRENDS : FOR LEADERS IN EDUCATION & TRAINING.
CALL NO : LB_1028.3 T41

FISL: P.EDUC
TITLE NAME: THEORY INTO PRACTICE
CALL NO : L____11.T49

FISL: P.EDUC
TITLE NAME: THE TIMES, LONDON. EDUCATIONAL SUPPLEMENT.
CALL NO : L____11.T5

FISL: P.EDUC
TITLE NAME: TODAY'S EDUCATION. COP.1
CALL NO : L____13.N2

FISL: P.EDUC
TITLE NAME: URBAN EDUCATION
CALL NO : LC_5101.U68

FISL: P.EDUC
TITLE NAME: VOCATIONAL ASPECT OF EDUCATION
CALL NO : LC_1041.V63

FISL: P.EDUC
TITLE NAME: VOCATIONAL EDUCATION JOURNAL. COP.1
CALL NO : LC_1041.A55

FISL: P.EDUC
TITLE NAME: YOUNG CHILDREN
CALL NO : LB_1140.J67

FISL: P.EDUC 60-00070
TITLE NAME: OREGON SCHOOL STUDY COUNCIL. BULLETIN.
CALL NO : LB____5.077

FISL: P.EDUC 60-00070
TITLE NAME: OREGON SCHOOL STUDY COUNCIL. MEMBERSHIP..
CALL NO : SEE COMMENTS

FISL: P.EDUC 60-00070
TITLE NAME: OSSC REPORT.
CALL NO : LB____5.0773

FISL: P.EDUC 64-01971
TITLE NAME: READING IMPROVEMENT
CALL NO : LB_1632.R4

FISL: P.EDUC 65 00761
TITLE NAME: COMMUNICATOR.
CALL NO : LB_2822.5 N2

FISL: P.EDUC 65-00761
TITLE NAME: NATIONAL ASSOCIATION OF ELEMENTARY SCHOOL PRINCIPALS. INSTITUTIONAL MEMBERSHIP
CALL NO : SEE COMMENTS

* FISL: P.EDUC 65-00761
TITLE NAME: PRINCIPAL.
CALL NO : L____11.N24

* FISL: P.EDUC 66-00105
TITLE NAME: EDUCATIONAL THEORY
CALL NO : L____11.E37

* FISL: P.EDUC 67-02305
TITLE NAME: CHRONICLE OF HIGHER EDUCATION.
CALL NO : LB_2300.C35

* FISL: P.EDUC 68-00256
TITLE NAME: COLLEGE STUDENT JOURNAL
CALL NO : LB_3602.C6

* FISL: P.EDUC 69-00056
TITLE NAME: PHI DELTA KAPPAN. (BINDING COPY)
CALL NO : LJ__121.P53

FISL: P.EDUC 69-00102
TITLE NAME: PHI DELTA KAPPAN. COP.1
CALL NO : LJ__121.P53

FISL: P.EDUC 69-01798
TITLE NAME: CURRENT ISSUES IN EDUCATION
CALL NO : LB____5.C8

FISL: P.EDUC 69-01798
TITLE NAME: INSIGHTS
CALL NO : L____11.I55

* FISL: P.EDUC 70-00259
TITLE NAME: MEASUREMENT AND EVALUATION IN COUNSELING AND DEVELOPMENT
CALL NO : LB_1027.5 M381

FISL: P.EDUC 70-01820
TITLE NAME: CHRONICLE OF HIGHER EDUCATION. MICROFILM ED.
CALL NO : LB_2300.C35

FISL: P.EDUC 70-02071
TITLE NAME: INDUSTRIAL EDUCATION. (MICROFICHE)
CALL NO : T____61.I5

FISL: P.EDUC 71-04220
TITLE NAME: JOURNAL OF COUNSELING AND DEVELOPMENT : JCD.
CALL NO : LC_1041.V6

FISL: P.EDUC 71-04537
TITLE NAME: COUNSELOR EDUCATION AND SUPERVISION
CALL NO : LB_2343.C685

FISL: P.EDUC 71-04636
TITLE NAME: JOURNAL OF COLLEGE STUDENT DEVELOPMENT
CALL NO : LB_2343.J6

FISL: P.EDUC 72-01208
TITLE NAME: ELEMENTARY SCHOOL GUIDANCE AND COUNSELING.
CALL NO : LB_1027.5 E5

FISL: P.EDUC 74-01269
TITLE NAME: MOMENTUM.
CALL NO : LC__461.M6

FISL: P.EDUC 74-01269
TITLE NAME: NATIONAL CATHOLIC EDUCATIONAL ASSOCIATION. LIBRARY MEMBERSHIP
CALL NO : SEE COMMENTS

FISL: P.EDUC 75-01306
TITLE NAME: DELTA KAPPA GAMMA BULLETIN
CALL NO : L____11.D4

FISL: P.EDUC 75-02227
TITLE NAME: EDUCATIONAL STUDIES
CALL NO : L____11.E354

FISL: P.EDUC 76-00045
TITLE NAME: CONTEMPORARY EDUCATION
CALL NO : L____11.T37

FISL: P.EDUC 77-01234
TITLE NAME: UNIVERSITAS.
CALL NO : LB_2300.U58

FISL: P.EDUC 78-10031
TITLE NAME: THE JOURNAL OF VOCATIONAL EDUCATION RESEARCH.
CALL NO : LC_1041.J68

FISL: P.EDUC 79-11146
TITLE NAME: COLLEGE TEACHING
CALL NO : LB_1778.I5

FISL: P.EDUC 81-12946
TITLE NAME: JOURNAL OF INSTRUCTIONAL PSYCHOLOGY
CALL NO : LB_1051.J63

FISL: P.EDUC 81-12960
TITLE NAME: SCHOOL PSYCHOLOGY REVIEW
CALL NO : LB_1051.S373

FISL: P.EDUC 81-12965
TITLE NAME: JOURNAL OF MORAL EDUCATION
CALL NO : LC__268.J67

FISL: P.EDUC 81-12978
TITLE NAME: URBAN REVIEW.
CALL NO : LC_5101.U7

FISL: P.EDUC 81-13256
TITLE NAME: RESEARCH IN HIGHER EDUCATION
CALL NO : LB_2331.63 R47

FISL: P.EDUC 81-13257
TITLE NAME: AMERICAN SECONDARY EDUCATION
CALL NO : L____11.A53

FISL: P.EDUC 81-13269
TITLE NAME: EDUCATIONAL PSYCHOLOGIST
CALL NO : LB_1051.E35

FISL: P.EDUC 81-13282
TITLE NAME: CONTEMPORARY EDUCATIONAL PSYCHOLOGY
CALL NO : LB_1051.C678

FISL: P.EDUC 82-10858
TITLE NAME: COMPUTING TEACHER
CALL NO : LB_1028.5 C5741

FISL: P.EDUC 84-03336
TITLE NAME: REMEDIAL AND SPECIAL EDUCATION : RASE.
CALL NO : LB_1029.R4 R41

FISL: P.EDUC 84-04726
TITLE NAME: JOURNAL OF READING. (BINDING COPY)
CALL NO : LB_1050.J6

FISL: P.EDUC 84-06659
TITLE NAME: EDUCATION DIGEST. COP.2 (MICROFILM)
CALL NO : L____11.E282 COP.2

FISL: P.EDUC 84-12131
TITLE NAME: PHI DELTA KAPPAN. COP.2 (MICROFILM)
CALL NO : LJ__121.P53 COP.2

FISL: P.EDUC 84-12135
TITLE NAME: JOURNAL OF NUTRITION EDUCATION. COP.2 (MICROFILM)
CALL NO : LB_1587.N8 J7 COP.2

FISL: P.EDUC 85-07930
TITLE NAME: CURRICULUM REVIEW
CALL NO : Z__1035.A1 C18

FISL: P.EDUC 85-07932
TITLE NAME: COMPUTERS IN THE SCHOOLS.
CALL NO : LB_1028.5 C57321

FISL: P.EDUC 85-15739
TITLE NAME: THE TECHNOLOGY TEACHER. (MICROFICHE)
CALL NO : TT__161.I49

FISL: P.EDUC 86-00667
TITLE NAME: CURRENT HEALTH 2.
CALL NO : RA__773.C8

FISL: P.EDUC 86-01677
TITLE NAME: VOCATIONAL EDUCATION JOURNAL. COP.2 (MICROFILM)
CALL NO : LC_1041.A55 COP.2

FISL: P.EDUC 86-03873
TITLE NAME: JOURNAL OF INDUSTRIAL TEACHER EDUCATION. (BINDING COPY)
CALL NO : TT__161.J6

FISL: P.EDUC 87-08944
TITLE NAME: JOURNAL OF EDUCATIONAL TECHNOLOGY SYSTEMS
CALL NO : LB_1028.3 J68

FISL: P.EDUC 87-08945
TITLE NAME: THE JOURNAL OF THE COLLEGE & UNIVERSITY PERSONNEL ASSOCIATION.
CALL NO : LB_2342.7 J681

FISL: P.EDUC 87-08946
TITLE NAME: JOURNAL OF EDUCATION FINANCE
CALL NO : LB_2825.J68

FISL: P.EDUC 87-08947
TITLE NAME: JOURNAL OF ECONOMIC EDUCATION
CALL NO : H____62.5 U5 J6

FISL: P.EDUC 87-08948
TITLE NAME: JOURNAL OF AUTISM AND DEVELOPMENTAL DISORDERS
CALL NO : RJ__499.A1 J58

FISL: P.EDUC 87-08949
TITLE NAME: JOURNAL OF LAW & EDUCATION.
CALL NO : K____10.0873

FISL: P.EDUC 87-08965
TITLE NAME: ASSOCIATION FOR PERSONS WITH SEVERE HANDICAPS. MEMBERSHIP
CALL NO : SEE COMMENTS

FISL: P.EDUC 87-08965
TITLE NAME: JOURNAL OF THE ASSOCIATION FOR PERSONS WITH SEVERE HANDICAPS
CALL NO : LC_4812.J681

FISL: P.EDUC 87-08966
TITLE NAME: INTERNATIONAL JOURNAL OF EARLY CHILDHOOD
CALL NO : LB_1140.A1 I56

FISL: P.EDUC 87-08967
TITLE NAME: EVALUATION REVIEW
CALL NO : HM____1.E8

FISL: P.EDUC 87-08971
TITLE NAME: EDUCATIONAL RESEARCHER
CALL NO : LB_1028.E32

FISL: P.EDUC 87-08972
TITLE NAME: EDUCATIONAL EVALUATION AND POLICY ANALYSIS.
CALL NO : LB_1028.E29

FISL: P.EDUC 87-08975
TITLE NAME: CHILDREN'S LITERATURE IN EDUCATION
CALL NO : Z__1037.A1 C51

FISL: P.EDUC 87-08976
TITLE NAME: BILINGUAL REVIEW.
CALL NO : ORDERED 12/19/86

FISL: P.EDUC 87-08977
TITLE NAME: CURRICULUM INQUIRY.
CALL NO : LB_1570.C96

FISL: P.EDUC 87-08979
TITLE NAME: ANTHROPOLOGY AND EDUCATION QUARTERLY
CALL NO : LB___45.C67A

FISL: P.EDUC 87-09058
TITLE NAME: JOURNAL OF DEVELOPMENTAL EDUCATION
CALL NO : LB_1029.R4 J641

FISL: P.EDUC 87-09060
TITLE NAME: JOURNAL OF COLLEGE ADMISSIONS
CALL NO : LB_2351.2 J68

FISL: P.EDUC 87-11926
TITLE NAME: OREGON SCHOOLS.
CALL NO : LA__352.0771

FISL: P.ELEC
TITLE NAME: AES : JOURNAL OF THE AUDIO ENGINEERING SOCIETY. AUDIO/ACOUSTICS/APPLICATIONS.
CALL NO : TK_6540.A92