8-1997

The Impact of Sex and Gender-Role Orientation on Student Evaluations of Professor Competence in Counselor Education and Counseling Psychology

Suzanne M. Hobson
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Counselor Education Commons, Gender and Sexuality Commons, and the Higher Education Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1631

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE IMPACT OF SEX AND GENDER-ROLE ORIENTATION ON
STUDENT EVALUATIONS OF PROFESSOR COMPETENCE
IN COUNSELOR EDUCATION AND
COUNSELING PSYCHOLOGY

by

Suzanne M. Hobson

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Counselor Education
and Counseling Psychology

Western Michigan University
Kalamazoo, Michigan
August 1997
THE IMPACT OF SEX AND GENDER-ROLE ORIENTATION ON STUDENT EVALUATIONS OF PROFESSOR COMPETENCE IN COUNSELOR EDUCATION AND COUNSELING PSYCHOLOGY

Suzanne M. Hobson, Ed.D.
Western Michigan University, 1997

The focus of this study was on the potential impact of sex and gender-role orientation on one form of evaluation within higher education. Specifically, this study investigated sex and gender-role orientation as they relate to graduate student end-of-course evaluations of professors in the Counselor Education and the Counseling Psychology fields.

Students enrolled in graduate courses in counselor education or counseling psychology at a large university in the Midwest completed the Instructional Development and Effectiveness Assessment (IDEA) end-of-course rating form, a modified version the Bem Sex-Role Inventory (BSRI) short form, and a student questionnaire. Professors also participated by completing the IDEA Faculty Information Form and a modified version of the BSRI. The completion of these instruments allowed for the collection of data regarding the three independent variables in this study (sex of student, sex of professor, and gender-role orientation of professor as perceived by students) and the three dependent variables used for analysis (rating of self-reported progress on 10 separate learning objectives, rating of the instructor, and rating of the course). The data
were analyzed for main effects and interaction effects using analyses of variance.

No evidence for interaction effects was found. The results yielded by these analyses failed to provide consistent evidence for a main effect of student sex. The results yielded by these analyses provided some evidence for a main effect of professor sex when including only this variable in the model. However, when professor gender-role orientation as perceived by students was also included in the model, the analyses failed to yield consistent evidence for a main effect of student sex. Finally, the results did yield consistent evidence for a main effect of professor gender-role orientation and suggested that this variable explained a significantly greater proportion of variability in the student ratings of professors than did student sex or professor sex.

These results may be useful to faculty members and administrators when considering issues or complaints of sex bias within student evaluations. They may also provide useful information about professorial qualities and styles appreciated by students within counselor education and counseling psychology.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
DEDICATION

This dissertation is dedicated to Mr. Dean Sousanis in appreciation of all that he offers. His career as a high school teacher has left an indelible mark on my life and has no doubt similarly influenced the lives of many others.

As a teacher,

a mentor,

a coach,

a friend,

an advocate,

and a colleague,

Dean Sousanis has given me reason to hope and a vision of purpose in my life. His involvement as a professional educator epitomizes the caring, commitment, and competence I hope to bring to my own work. His dedication to and immersion in efforts to help young people are qualities I will strive to instill in those I teach. It is in loving recognition of his continuing influence in the lives of so many that I dedicate this dissertation to him.
ACKNOWLEDGMENTS

It is common wisdom that those who achieve do so only with the assistance and support of many others and that any achievement is not an individual accomplishment but a collaborative one. Among those who have assisted and supported me through the dissertation process have been faculty, friends and family.

I would first like to acknowledge and thank the members of my doctoral committee: Dr. Edward Trembley, Dr. Suzanne Hedstrom, and Dr. Susan Caringella-MacDonald. Their ability and willingness to work collaboratively on my behalf was essential to my feeling supported and validated throughout my doctoral program. I would also like to thank Dr. Suzanne Hedstrom and Dr. Donna Talbot for their generosity of time and spirit in mentoring me along the way. In addition, I would like to acknowledge the contributions of Dr. Bradley Huitema, Dr. John Geisler, Ms. Julie Scott, and Ms. Amalia Meier for their assistance with the design of the study and the analysis of the data.

I would also like to acknowledge three particular family members: my mother, Virginia Gould Hobson; my maternal grandmother, Beverly Howland Gould; and my paternal grandmother, Georgia Luella Hobson. My attention to and interest in issues regarding women and work was largely inspired by the lessons taught by them. Thank you! Finally, thankyou also to all of my family and friends who continuously provided support and encouragement. It is only because of you that I could complete and actually enjoy this entire process!

Suzanne M. Hobson

ii
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** ......................................................................................... ii  
**LIST OF TABLES** ...................................................................................................... ix  
**LIST OF FIGURES** ...................................................................................................... xi  

## CHAPTER

**I. INTRODUCTION** ................................................................. 1  
Introduction and Statement of the Problem. ............................................. 1  
Purpose of and Need for the Study ......................................................... 3  
Definitions ................................................................................................. 4  
Research Questions ................................................................................... 7  
Hypotheses ................................................................................................ 7  
Scope and Delimitations .......................................................................... 9  
Overview of Methodology ....................................................................... 11  
Outline of Dissertation ............................................................................ 12  

**II. REVIEW OF RELATED LITERATURE** ................................................. 13  
Historical Background ............................................................................ 13  
Female Access to Employment ......................................................... 13  
Anti-Discrimination Legislation ....................................................... 14  
Current Situation .................................................................................... 17  
The Glass Ceiling ..................................................................................... 19

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
## Table of Contents—Continued

### CHAPTER

**Sex-Role Stereotypes** ............................................................. 22

**Faculty Evaluations in Higher Education** ................................. 26

**Criteria for Tenure** ............................................................. 26

**Evaluation of Teaching Effectiveness** ......................................... 28

**Student Evaluations of Teaching Effectiveness** ......................... 31

**History and Development of Student Evaluations** ..................... 31

**Measurement of Student Evaluations** ....................................... 34

**Global Versus Specific Measures** ............................................ 34

**Student Evaluation Instruments** ............................................. 40

- **Instructional Development and Effectiveness Assessment (IDEA)** ................................................................. 41
- **Student Instructional Report (SIR)** ........................................ 42
- **Student's Evaluation of Educational Quality (SEEQ)** .............. 43
- **Endeavor Instructional Rating System (EIRS)** ....................... 45
- **Instructor and Course Evaluation System (ICES)** .................. 45
- **Student Instructional Rating System (SIRS)** ....................... 45
- **Instructional Assessment System (IAS)** ............................... 46

**Research on the Accuracy of Student Evaluations** .................... 46

**Reliability of Student Evaluations** ......................................... 47

**Validity of Student Evaluations** ............................................ 50
# Table of Contents—Continued

**CHAPTER**

- Correlations With Student Learning ........................................... 51
- Correlations With Other Ratings of Teacher Effectiveness .... 53
- Research on Potential Bias ................................................................. 59
- Student Characteristics ................................................................. 59
- Course Characteristics ............................................................... 61
- Instructor Characteristics ............................................................ 65
- Sex, Gender-Role Orientation, and Student Evaluations ............. 71

**III. DESIGN** ................................................................. 78

- Introduction .............................................................................................. 78
- Population and Sample .......................................................................... 78
- Instrumentation ....................................................................................... 79
  - Introduction ....................................................................................... 79
  - Instructional Development and Effectiveness Assessment System .. 80
  - Modified Bem Sex-Role Inventory ................................................... 82
  - Student Questionnaire ...................................................................... 84
  - IDEA Faculty Information ................................................................ 84
  - Preparation of Instrumentation ......................................................... 85
- Methodology ............................................................................................ 85
  - Preparation ....................................................................................... 85
  - Data Collection ................................................................................ 86

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—Continued

CHAPTER

Coding of Response Sheets .............................................................. 88
Scoring ............................................................................................. 89
Data Analysis ..................................................................................... 89
Limitations ............................................................................................. 94
Ethical Issues ........................................................................................... 95

IV. FINDINGS .................................................................................................. 96
Introduction ............................................................................................. 96
Description of Participating Students ..................................................... 97
Description of Participating Professors ................................................. 97
Data on Reduction of Sample for Statistical Analyses ....................... 101
Description of Classes Surveyed .......................................................... 103
Tests of the Assumptions ........................................................................ 105
Results ................................................................................................. 106
Introduction ..................................................................................... 106
Restatement of Research Questions and Hypotheses ................... 107
Results Relevant to Research Question 1 ...................................... 107
Results Relevant to Research Question 2 ...................................... 113
Results Relevant to Research Question 3 ...................................... 117
Results Relevant to Research Question 4 ...................................... 125
Results Relevant to Research Question 5 ...................................... 126
# Table of Contents—Continued

## CHAPTER

Summary of Results ................................................................. 128

V. DISCUSSION OF FINDINGS: IMPLICATIONS AND EXTENSIONS. 130

Introduction ............................................................................. 130

Purpose and Design of the Study ............................................. 130

Interpretation and Discussion of Results ............................. 132

Results Relevant to Research Question 1 ............................. 132

Summary of Results ................................................................ 133

Discussion ............................................................................. 134

Results Relevant to Research Question 2 ............................. 134

Summary of Results ................................................................ 134

Discussion ............................................................................. 135

Results Relevant to Research Question 3 ............................. 136

Summary of Results ................................................................ 137

Discussion ............................................................................. 138

Results Relevant to Research Question 4 ............................. 141

Summary of Results ................................................................ 142

Discussion ............................................................................. 142

Results Relevant to Research Question 5 ............................. 142

Summary of Results ................................................................ 143

Discussion ............................................................................. 143
CHAPTER

Limitations ................................................................. 144
Suggestions for Future Research ................................. 147
Summary and Conclusions ........................................... 149

APPENDICES

A. Rules for Teachers, 1872 ........................................... 150
B. A Narrow Band of Acceptable Behaviors for Women .... 152
C. Letter of Permission to Utilize the Instructional Development and Effectiveness Assessment (IDEA) Instrument ....... 154
D. HSIRB Letter of Approval ........................................... 156
E. Letter of Permission to Utilize Bem Sex-Role Inventory (BSRI) Instrument ........................................... 158
F. Student Questionnaire ................................................. 160
G. Oral Recruitment Script ............................................. 162
H. Informed Consent Form ............................................. 164
I. Results of 2 Factor ANOVAs ...................................... 166
J. Results of 1 Factor ANOVAs ...................................... 179

BIBLIOGRAPHY .............................................................. 192

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
LIST OF TABLES

1. Correlation of Student Ratings: Teacher vs. Course Characteristics .......... 50
2. Data Output Columns .................................................................................. 90
3. Comparison of Total Student Participants and Population ....................... 98
4. Total Participating Professors by Sex .......................................................... 98
5. Sex of Student and Professor Participants .................................................. 99
6. Gender-Role Orientation of Participating Professors: A Comparison of Student Descriptions and Professor Self-Report ............................................ 100
7. Original Cell Counts and Cell Counts After Reduction: 2 Factor ANOVAs With Professor Sex and Professor GRO in Model .............................. 102
8. Original Cell Counts and Cell Counts After Reduction: 2 Factor ANOVAs With Professor Sex and Student Sex in Model ....................................................... 103
9. Original Cell Counts and Cell Counts After Reduction: 2 Factor ANOVAs With Student Sex and Professor GRO in Model ....................................................... 104
10. Total Participating Classes and Total Participating Students in Classes Taught by Male and Female Professors .............................................................. 104
11. Restatement of Research Questions and Hypotheses ................................ 108
12. Test of Model Significance for 2 Factor ANOVA Including Student Sex and Professor Sex in Model ................................................................. 110
13. Test of Model Significance for 2 Factor ANOVA Including Student Sex and Professor Gender-Role Orientation in Model ....................................................... 111
14. ANOVA Summary Table for 2 Factor ANOVA Including Student Sex and Professor Gender-Role Orientation in Model ....................................................... 111
15. Mean Ratings of Professor Competence by Male and Female Students in 2 Factor ANOVAs including Student Sex and Professor GRO in Model ...... 112
List of Tables—Continued

16. Test of Model Significance for 2 Factor ANOVA Including Professor Sex and Student Sex in Model .......... 114

17. Test of Model Significance for 2 Factor ANOVA Including Professor Sex and Professor Gender-Role Orientation in Model .......... 115

18. ANOVA Summary Table for 2 Factor ANOVA Including Professor Sex and Professor Gender-Role Orientation in Model .......... 115

19. Mean Ratings of Professor Competence for Male and Female Professors in 2 Factor ANOVAs Including Professor Sex and Professor GRO in Model .... 116

20. Test of Model Significance for 2 Factor ANOVA Including Professor Gender-Role Orientation and Student Sex in Model .......... 119

21. ANOVA Summary Table for 2 Factor ANOVA Including Professor Gender-Role Orientation and Student Sex in Model .......... 119

22. Mean Ratings of Professor Competence for Professors in 2 Factor ANOVAs Including Professor GRO and Student Sex in Model ........ 120

23. Probability of Statistically Significant Differences Between Ratings of Professor Competence According to Professor Gender-Role Orientation .......... 121

24. Test of Model Significance for 2 Factor ANOVA Including Professor Gender-Role Orientation and Professor Sex in Model .......... 122

25. ANOVA Summary Table for 2 Factor ANOVA Including Professor Sex and Professor Gender-Role Orientation in Model .......... 123

26. Mean Ratings of Professor Competence for Professors in 2 Factor ANOVAs Including Professor Sex and Professor GRO in Model ........ 124

27. Probability of Statistically Significant Differences Between Ratings of Professor Competence According to Professor Gender-Role Orientation .......... 125

28. Proportion of Variability in Student Ratings of Professor Competence Explained by Each Independent Variable .......... 127

29. Summary of Findings .......... 129

x
LIST OF FIGURES

1. Criteria Used in Faculty Evaluations for Tenure and Promotion Decisions . . . 28
CHAPTER I

INTRODUCTION

Introduction and Statement of the Problem

The evaluation of professors in higher education has been extensively researched. This research has consisted of an identification of the various responsibilities of professors which should be considered for evaluation, an examination of the characteristics of effective teaching, an investigation of methods of obtaining reliable and valid ratings of professor competence, and an exploration of the appropriate uses of each type of evaluation. Despite the extensive nature of previous research, the question of how each of the above areas of study are affected by sex has continued to be raised. Research on sex has focused on whether discrimination against females has persisted within higher education and, if so, on the dynamics underlying the discrimination.

Sandler and Hall (1986) argued convincingly that discrimination against women continues to exist within higher education and that it targets both students and faculty members. They suggested that female students and faculty members alike face a "chilly climate" on campuses across the nation (1986, p. 2). Researchers have documented the underrepresentation of females in the higher academic ranks, such as Full Professor, or administrative ranks, such as Dean or Provost (Hall, 1982; Harris, 1975; Sandler & Hall, 1986; York, Henley, & Gamble, 1985). Other studies have found that females are
underrepresented at the more prestigious universities, tend to be clustered in the small universities and community colleges, and are often employed only on a part-time, non-tenure tracked basis (Feldt, 1990; Maitland, 1990; Simeone, 1987). Researchers have also offered evidence that females receive significantly lower salaries than males with comparable experience (Maitland, 1990; Stroh, Brett, & Reilly, 1992; Williams & McCullers, 1983; York, Henley & Gamble, 1985).

Bacchi (1993) argued that such underrepresentation of females is indicative of a lack of advancement opportunities for women and can be attributed to a phenomenon similar to the glass ceiling effect first observed within the field of business management.

Most people have heard of the 'glass ceiling'—that impenetrable barrier facing women who aspire to top management positions. There is another domain which has proved equally resistant to women's efforts to enter—the higher echelons of the academic hierarchy. Here women have encountered a brick wall. (Bacchi, 1993, p. 36)

Other researchers have suggested that this effect is a direct result of gender-role stereotypes (Bray & Howard, 1980; Maitland, 1990; Martin, 1984; Sidanius & Crane, 1989). According to Etaugh and Riley (1983), "considerable evidence exists that women are often judged to be less competent than men with equal qualifications and accomplishments" (p. 943). Research has found that even when assuming that all else is equal, women are often evaluated as being less competent than men. This would seem to support the notion that gender-role stereotypes are contributing to the devaluation of women in work settings and therefore to the glass ceiling effect, or -- in Bacchi's words -- the construction and maintenance of the brick wall.
Purpose of and Need for the Study

This study further examined the potential impact of sex and gender-role stereotypes on one form of evaluation within higher education. Specifically, this study investigated sex and gender-role orientation as they relate to graduate student evaluations of professors in the Counselor Education and Counseling Psychology fields.

According to Centra (1979), student evaluations are used by most institutions as the primary summative measure of teaching effectiveness. Although a considerable amount of research has studied these variables (sex and gender-role orientation) in conjunction with student evaluations of professor competence (Basow & Silberg, 1987; Bennett, 1982; Elmore & LaPointe, 1974; Harris, 1975; Harris, 1976; Kanekar, 1990; Kaschak, 1981; Kierstead, D'Agostino, & Dill, 1988; Martin, 1984; Sidanius & Crane, 1989), much of this research is somewhat dated and it has all focused on the undergraduate level. Finally, none of the research found has addressed these variables exclusively within the fields of counselor education and counseling psychology.

This study was expected to determine whether student evaluations of professors continue to be biased by sex and/or gender-role orientation at the present time. This study also sought to provide information regarding the existence and operation of these dynamics at the graduate school level by focusing on a graduate program in a department of Counselor Education and Counseling Psychology.

Furthermore, the focus of the majority of past studies has been on researching whether there is a difference in student ratings across sex. Only a handful of studies have
explored the relationship between gender-role orientation and student ratings (Bray & Howard, 1980; Harris, 1975; Harris, 1976; Wheeless & Potorti, 1989). By including gender-role orientation as a variable of interest, this study purported to be helpful in clarifying the relative importance of sex and gender-role orientation in order to better understand the dynamics underlying the glass ceiling effect.

Studying the nature of the association between sex, gender-role orientation, and student ratings of professor competence may have important implications for our ability to effectively use student ratings as a summative measure of teaching effectiveness. In addition to providing information which may enhance the quality of tenure and promotion decisions at universities, this research was expected to contribute to the understanding of the career experience of women in male-dominated occupations.

Definitions

The following definitions are provided to clarify the meaning and use of certain terms used in this study:

**Sex:** Sex is defined as an individual's biological sex. Students and professors who participated in this study were categorized as either male or female.

**Gender-role orientation:** Gender-role orientation is defined as "the degree to which [individuals] identify with an array of gender-typed attributes" (Bieger, 1985, p. 51). Within the literature, the terms 'gender-role orientation' and 'sex-role orientation' and 'sex-role self-concept' have been used interchangeably. With the exception of direct quotations, this study relied exclusively on the term 'gender-role orientation.' For the
purposes of this study, gender-role orientation was determined by the administration of the Bern Sex-Role Inventory (BSRI) (Bem, 1978). Professors were thereby categorized as masculine, feminine, androgynous, or undifferentiated based on the perceptions reported by students on the modified version of the BSRI.

Individuals who have a masculine or a feminine gender-role orientation are described as sex-typed: "highly attuned to cultural definitions of sex-appropriate behavior and ...[using] such definitions as the ideal standard against which" to evaluate their behavior (Bem, 1981, p. 5). Individuals who have an androgynous or undifferentiated gender-role orientation are not said to be sex-typed.

**Masculine:** For the purposes of this study, the word 'masculine' will refer to one of four possible gender-role orientations as measured by the BSRI. A masculine sex-role orientation reflects "an instrumental (cognitive, active) role that includes a goal-directed, task-oriented, and independent perspective" (Borders and Fong, 1984, p. 59).

**Feminine:** For the purposes of this study, the word 'feminine' will refer to one of four possible gender-role orientations as measured by the BSRI. A feminine sex-role describes "an expressive (nurturant, receptive) role characterized by an affective, supportive, other-focused perspective" (Borders and Fong, 1984, p. 59).

**Androgynous:** For the purposes of this study, the word 'androgynous' will refer to one of four possible gender-role orientations as measured by the BSRI. Androgynous individuals are those who possess a high number of both masculine and feminine characteristics.

**Undifferentiated:** For the purposes of this study, the word 'undifferentiated' will
refer to one of four possible gender-role orientations as measured by the BSRI. Undifferentiated individuals are those who possess a low number of both masculine and feminine characteristics.

Professor competence: In this study, professor competence refers to an individual professor's objective quality as an instructor. Because there are no universally accepted criteria with which to evaluate this, however, the actual objective quality of instructor is largely unmeasureable (Marsh, 1982).

Student evaluations of professor competence: Student evaluations of professor competence in this study are defined by the mean rating assigned to a given professor on global measures of instructor competence obtained using end-of-course surveys. Only global measures were used to determine student evaluations of professor competence because researchers have suggested that only they, as opposed to measures of specific behaviors or styles, be used for summative purposes including tenure and promotion decisions (Abrami & d'Apollonia, 1990; Centra, 1979, 1993). Throughout this dissertation, references to ratings of professor competence will be used interchangeably with "ratings of instructor" and "ratings of teacher."

The instrument utilized in this study to obtain student ratings of professor competence was the Instructional Development and Effectiveness Assessment (IDEA) System (Hoyt & Cashin, 1977). Additional information regarding which items are considered global measures and how the score were computed are included in the instrumentation section in Chapter III.
Research Questions

This study involved the collection and analysis of data designed to address the following research questions:

1. Is there a significant difference between ratings of professor competence provided by male and female students?

2. Is there a significant difference between ratings received by male professors and female professors?

3. Is there a significant difference among the ratings received by professors who are perceived as masculine, feminine, androgynous, or undifferentiated by students?

4. Are there interaction effects between student sex and professor sex, between professor sex and gender-role orientation, or between student sex and professor gender-role orientation?

5. Does sex or gender-role orientation explain a greater proportion of the variability in the student rating of professor competence?

Hypotheses

With regard to the first research question, it was hypothesized that there would be no significant difference between ratings of professor competence by male and female students.

In response to the second research question, male professors were expected to receive higher ratings of competence from students than female professors.
With regard to the third research question, it was hypothesized that professors who are perceived as masculine or androgynous would receive higher ratings from students than those who are perceived as feminine or undifferentiated.

In response to the fourth research question, interaction effects were anticipated. Sex was expected to interact with gender-role orientation. It was hypothesized that male professors would tend to be seen (by self-report or student description) as masculine or androgynous and that female professors would tend to be seen (again, by self-report or student description) as androgynous or feminine. Interaction effects between student sex and professor sex were anticipated. It was hypothesized that female professors would receive higher ratings from female students and that male professors would receive higher ratings from male students.

With respect to the fifth and final research question, it was hypothesized that gender-role orientation would explain a greater proportion of the variability in the student ratings of professor competence than would the sex of the professor or of the student. If it could be shown that this is indeed the case, this study would provide evidence to argue against the existence of simple sex bias against women in higher education and instead suggest more complicated dynamics. Such a finding might also have explanatory power in illuminating the reasons for the inconsistent findings with regard to sex bias in the workplace.

The research questions, research hypotheses, and associated null hypotheses are presented in Table 11 located in Chapter IV.
Scope and Delimitations

The scope of this study was necessarily limited by its focus on student evaluations of a professor's teaching. Universities traditionally consider three different types of responsibilities when doing a summative evaluation of a professor (Centra, 1979; Centra, 1993): teaching, research and publication, and service (such as work on committees or on community projects). Teaching, which was the focus of this study, constitutes only one of these three areas. Lin, McKeachie, and Tucker (1984) questioned the actual importance of teaching evaluations in the tenure and promotion decisions made by universities: "Most colleges and universities attach great importance to teaching ability in criteria for promotion, yet faculty members everywhere believe that the formal criteria claim more weight for teaching than is actually accorded it" (p. 583). Other researchers have suggested that the extent to which a university may state the importance of teaching but weight more heavily research and publication represents a systemic devaluation of women by devaluing teaching. Martin (1984) stated that the university "reward structure places greater value on traditional male activities [such as research and publication and administration] than on female activities [such as teaching]" (p. 483).

A second limitation of this study emerges from its emphasis on student ratings of instructors. Although researchers have documented a general agreement between student ratings and ratings by colleagues (Centra, 1975; Centra & Bonesteel, 1990; Elmore & LaPointe, 1974; Feldman, 1988; Marsh, 1982), other researchers have suggested that there may be a difference between these evaluations and those made by
the actual administrators who make the tenure and promotion decisions (Coufal & Hines, 1976; Harris, 1975; Harris, 1976). Coufal and Hines called attention to this discrepancy: "In practice, faculty evaluations are often based on meager information, on the 'gut feelings of deans and department chairmen [sic]'" (1976, p. 5). This study investigated the impact of sex and gender-role orientation only on student evaluations. Because sex becomes a more salient issue when less information is available (Sidanius & Crane, 1989), there may actually be more of a bias when evaluations of teaching effectiveness are done by administrators who have not spent the entire semester in a professor's class.

A third limitation of this study is that it specifically examined only sex and gender-role orientation as possible explanations for the lack of advancement of women within higher education. Researchers have offered three different theories to explain the 'glass ceiling effect' (Morrison & Von Glinow, 1990). According to Morrison and Von Glinow, the first theory states that women and other minorities have deficiencies which contribute to their inability to advance at a rate commensurate with males. These deficiencies might include a fear of success (Horner, 1972), an adversity to risk-taking, or lesser intellectual capacities. A second theory suggests that "discrimination by the majority population [is]... the major cause of the inequities" (Morrison & Von Glinow, 1990, p. 201). The idea of sex discrimination and sex-role stereotypes fits within this theory and was examined in this study. Finally, the third theory points to systemic barriers, such as tokenism (Kanter, 1977; Young, MacKenzie, & Sherif, 1980), demands for biculturalism, or a lack of mentors (Holt, 1981; Smith, 1982; York, Henley & Gamble, 1985) in order to explain women's lack of upward mobility. The scope of this
study investigated components of only one of these three theories.

Other limitations of this study are discussed in Chapter III and will be related to the methodology and scope of this study. The limitations of this study will then be revisited in Chapter V and used as the foundation for recommendations for future research.

Overview of Methodology

A sample of male and female graduate students completed a set of three instruments: (1) the Instructional Development and Effectiveness Assessment (IDEA) System to evaluate their professor, (2) a modified version of the Bem Sex-Role Inventory (BSRI) Short Form to describe their professor, and (3) a questionnaire designed by the researcher to provide additional data. This data consisted of demographic data allowing each participant to be categorized as a man or woman for use in later analysis and one item in which the student rated the overall quality of the course.

Professors completed two instruments: (1) the IDEA Faculty Information Form (an instrument to allow them to assign weights to reflect the importance of 10 student progress items on the IDEA instrument), and (2) the BSRI Short Form to describe themselves. Each professor was categorized in three ways for later analysis: (1) as a male or female; (2) as masculine, feminine, androgynous or undifferentiated as rated by the students; and (3) as masculine, feminine, androgynous or undifferentiated as reported by self.

The data were then analyzed using analyses of variance (ANOVAs) to identify
any main effects and/or interaction effects of student sex, professor sex, and professor gender-role orientation as rated by the students on each of the three measures of professor competence collected from the students.

Outline of Dissertation

Chapter II of this dissertation provides a review of related literature. This review includes sections on the historical background of women and work, on the glass ceiling, on gender-role stereotypes, and on the evaluation of professors in higher education. Chapter III will describe the design of the study. Sections are dedicated to the population studied, to the instrumentation which was utilized, and to methodological issues including data collection and data analysis. Chapter IV consists of the results obtained in the study. Chapter V includes a discussion of the results and their implications, attention to possible limitations of the study, and suggestions for future research.
CHAPTER II

REVIEW OF RELATED LITERATURE

Historical Background

Females have historically been discriminated against in the workplace (Hall, 1982; Holt, 1981; Kanter, 1977; Kessler-Harris, 1985). Although having finally gained access to male-dominated professions, women continue to struggle for equal treatment (Basow & Silberg, 1987; Morrison, White, & Van Velsor, 1992; Morrison & Von Glinow, 1990; Schuss, 1994; Sidanius & Crane, 1989). This section will summarize the legislative attempts to provide females with equal access to employment and to eliminate discrimination on the basis of sex. It will examine these issues as they pertain both to work in general and to work within higher education specifically.

Female Access to Employment

Over the years, women have struggled to gain equal access to employment in the United States. Women have historically been limited to working within traditionally female occupations such as clerical work (Kessler-Harris, 1985), and have had only limited access to male-dominated professions (Hall, 1982; Holt, 1981; Kanter, 1977; Kessler-Harris, 1985; Sandler & Hall, 1986). This strong division between "male occupations" and "female occupations" began to be questioned during the early 1900's.
Just prior to the emergence of this movement the 1890 census data revealed that only 17 percent of the work force was female. Of these working women, approximately 75 percent were single (Smith, 1979).

One factor which served as a catalyst for the entry of women (including those who were married) into the work force was the Great Depression (Kessler-Harris, 1985; Smith, 1979). The depression era marked a time of economic urgency and represented the beginning of a widespread societal movement to accept increasing numbers of women into the workforce. The financial difficulties faced by American families effectively debunked the myth that the husband could (even if he should) be the sole breadwinner in the family. As a result, the percentage of female workers who were married increased several percentage points to 35.

This movement gathered momentum as Rosie the Riveter (Burke, 1994) and the idea of a woman doing "a man's work" was popularized during World War II. "During the war years, women demonstrated their capacity to work at the same jobs and as effectively as men" (Kessler-Harris, 1985, p. 150).

Anti-Discrimination Legislation

The feminist movement that followed, demanding equal rights for women, resulted in an array of legislation. According to Kessler-Harris (1985), the seeds of this legislation can be traced back to the Fair Labor Standards Act in 1938. Kessler-Harris described this law as revolutionary because it was one of the earliest legal movements designed to minimize the differences between males and females in the work force. The
Fair Labor Standards Act sought to accomplish this by awarding to males the same protection with regard to a minimum wage and a limit on the number of hours one could work in a given week. Prior to this time, only females were viewed as needing special protection and this need was attributed to differences or weaknesses (Kessler-Harris, 1985).

Even with the Fair Labor Standards Act, though, sex discrimination continued to exist. Because it included "so-called protective labor laws" which limited the type of work women could do, legislated against working while pregnant, and prohibited working during certain hours (e.g., at night), the Fair Labor Standards Act became a "mechanism through which firms could legally refuse to hire women for certain jobs" (Barrett, 1979, pp. 54-55).

The Equal Pay Act, passed in 1963, represented the next legislative move against sex discrimination (Barrett, 1979; Kessler-Harris, 1985). This law specified that, once hired, males and females must be paid equally for equal work. Barrett (1979) explained that the legislation outlawed "separate pay scales for men and women for work requiring similar skills and performed under similar working conditions" (p. 55). Despite the fact that the intent of this legislation to eliminate sex discrimination in the workplace was often sabotaged by discrimination in hiring and promotion practices and served to maintain the occupational segregation by sex (Simeone, 1987), Barrett viewed this law as a "a watershed, a turning away from a legal system that facilitated and sanctioned discrimination against women to a legal environment that prohibits sex discrimination in employment" (1979, p. 55).
The loopholes in the Equal Pay Act were legally eliminated in 1964 with the passage of the Civil Rights Act. Title VII of this act prohibited "all forms of discrimination in employment, including hiring, firing, promotion, training, and fringe benefits...[and] established an enforcement agency, the Equal Rights Employment Opportunity Commission (EEOC)" (Barrett, 1979, p. 55). Interestingly, neither the Equal Pay Act nor Title VII of the Civil Rights Act applied to professionals within education (Barrett, 1979).

Therefore, women working in higher education were excluded from the equal rights afforded by these two pieces of legislation. It was eight years before legislation was passed which protected these women and those working within other areas of education from sex discrimination. Beginning in 1972, three laws were passed which specifically addressed the issue of sex discrimination in higher education. Title IX of the Higher Education Act, which was passed in 1972, prohibited discrimination against women and especially targeted the area of equal access to higher education for female students (Hall, 1982). By allowing increasing numbers of females to enter college and graduate school, Title IX resulted in women becoming increasingly qualified for professional positions traditionally held only by men. This resulted in less occupational segregation and the entrance of females into non-traditional fields such as the "hard" sciences. Also, as colleges and universities were challenged to become aware of and remedy teaching and administrative behaviors which might discriminate against female students, those women teaching in higher education also became more aware of how they experienced subtle or overt discrimination.
Thus, in an indirect way, Title IX had a definitive impact on women working in higher education. Barrett (1979) indicated that the Educational Amendments Act of 1972 and the Women's Educational Equity Act of 1974 addressed the issues regarding sex discrimination toward female professors directly. She stated that, in addition to allowing "women teachers to seek legal redress..., [they] nullified school board practices requiring pregnant teachers to take leave without pay...and opened the door to elimination of sex biases in school curricula and academic programs" (p. 56).

Indeed, women have more equal access to employment today than they have had in the past. An examination of the field of teaching, for instance, reveals vast changes. In 1872, female teachers were required to resign when they married or "engaged in unseemingly conduct" (Rules for Teachers, 1872; see Appendix A). At that time, the belief was that a woman's duties to her husband and family required all of her time and energy and that none should be devoted to gainful employment. The Educational Amendments Act and the Women's Educational Equity Act marked the legal end of such requirements.

**Current Situation**

Now, 56 percent of all women work, with approximately 58,501,000 women employed in the civilian labor force, 3,356,000 women reporting unemployment, and 42,528,000 women identifying themselves as not being in the labor force (Green & Becker, 1997). However, the legislation and anti-discrimination policies discussed above have not eradicated discrimination and bias against women in the workplace (Kessler-Harris,
1985; Sandler & Hall, 1986). Kessler-Harris (1979) lamented this fact as she noted that all of the legislation described above, when put together, still "have had little discernible impact for women as a whole" (p. 153). In 1986, Sandler and Hall found that, "despite many improvements, some things have not changed at all" within academe (1986, pp. 1-2). These authors cited a variety of statistics showing that female employees within academe tend to hold lower ranked, less esteemed and less lucrative positions than their male counterparts. These complaints appear to remain valid today. Green and Becker (1997) reported that women today represent 43.5 percent of all college and university teachers and that women today represent 38 percent of full-time college and university teachers. They also reported that the median weekly salary of full-time female college and university teachers is only 81 percent of the median weekly salary received by males working as college or university teachers. Furthermore, Reich and Nussbaum (1994) suggested that this may be an overestimate of women's earning power: "The gap between the earnings of women and men is smaller for weekly wages than for annual earnings" (p. 29).

Simeone (1987) offered insight into the reasons for the lack of change in spite of the many anti-discrimination and affirmative action laws now in place:

Because the vast majority of those in decision-making positions are male, systems are formed which conform more closely to the needs and experiences of academic men than women....One cannot absolve academia of sexist discrimination when it is manifested in the acts of individuals, for those acts are reinforced by the assumptions, policies, and practices of the institutions themselves. The research demonstrates clearly that on the whole, institutional systems have outcomes which are less favorable to women than to men. (p. 143)

One such outcome seems to be that, despite legally mandated equal access to
employment within higher education as well as other fields, females are advancing more slowly than men and tend to hold lower level positions and receive lesser salaries (Feldt, 1990; Hall, 1982; Harris, 1975; Maitland, 1990; Sandler & Hall, 1986; Simeone, 1987; York, Henley & Gamble, 1985). This phenomenon is often referred to as the glass ceiling effect.

The Glass Ceiling

A variety of attempts have been made to define the phenomenon known as the glass ceiling. Kanter (1977) wrote about the phenomenon before it had been labeled as such. She referred to workers who become stuck and find themselves at dead ends in a corporation. Reasons for such limited upward mobility included holding a job which never did offer much opportunity for advancement, failing to rise as a result of losing the opportunity to other competitors, and having further opportunities for advancement limited because of having "come through the wrong route" (Kanter, 1977, p. 138). This third type of dead end most closely resembles a glass ceiling.

These were generally the people who reflected whatever commitment the company had to improving opportunities for the disadvantaged—for women and minorities, for clerical and factory workers....There was just enough openness and an occasional example to keep the illusion that even people in low status jobs or without all of the right credentials [such as being male] could eventually work their way up, but the professional or management post that represented the way out of the low mobility job cluster was likely to come with its own very low ceiling (emphasis added). (p. 139)

Following Kanter's lead, the literature on women and minorities in the workplace began to address this issue. According to Morrison and Von Glinow (1990), the glass
ceiling represents a "concept popularized in the 1980's to describe a barrier so subtle that it is transparent, yet so strong that it prevents women and minorities from moving up the management hierarchy" (p. 200). Since that time, the idea of the glass ceiling has been broadened to encompass occupations existing outside of corporate America.

Schuss (1994) redefined the glass ceiling as any "career barrier that prevents qualified people from advancing and reaching their full potential within an organization" (p. 6). The organization may or may not exist within the business world. Furthermore, "reaching their full potential" does not necessarily refer to reaching a top management position. To make this point, Schuss quoted Joyce Miller, the executive director of the National Glass Ceiling Commission: "I'm equally concerned with women whose feet are stuck in the mud...Everyone wants to be president..., but whatever level they're at, they ought to be able to rise to whatever level they feel comfortable with" (cited in Schuss, 1994, p. 6).

The fact that female professors continue to struggle for equality in what has traditionally been a male-dominated field (Basow & Silberg, 1987) may therefore be considered indicative of a glass ceiling operating within university settings. According to Sandler and Hall (1986), "the difficulties that women face in the academic community are not that different from those faced by other professional women in the world at large" (p.17). Harris (1975) indicated that "the proportion of women [in university teaching] has traditionally been small, particularly in higher academic ranks and tenured positions" (p. 751). In addition to holding a low proportion of the tenured positions, females have also struggled to have their competence recognized with promotions to administrative
positions. In fact, Sandler and Hall found that "on the average, colleges and universities nationwide employ only 1.1 senior woman (dean and above) per institution" (1986, p. 14). Despite initially believing that they could advance without limitation according to their level of individual competence, female faculty members seem to be finding, in Kanter's terms, that their opportunities for advancement are limited by having "come through the wrong route" (Kanter, 1977, p. 138).

Young, MacKenzie, and Sherif (1980) suggested that this "wrong route" involves an interaction between sex and organizational demands for compliance. Token women, they found, are allowed to advance (gain tenure) on the condition that they subscribe to a belief system which sets them apart from other women and which suggests that their advancement was based on merit. These women serve to perpetuate the glass ceiling by denying the existence of sex discrimination.

Others, however, have pointed to the clear existence of such discrimination. Hall (1982) asserted that one reason for the lack of advancement of women is that the old saw that "a woman must be twice as good to get half as far as a man" still contains a core of truth: our society tends in many ways to value men more than it values women, and to assume that men's work and words are important, women's less so. (p. 4)

This suggests that there may be a variety of reasons for the relative lack of advancement of female professors. Hall suggested that universities require more work and ability from women in order to value them as highly as men. Young, et. al. (1980) suggested that a woman must accept a gender-stereotypic role of compliance in order to advance.
It seems, then, that there is a double standard. On the one hand, women are expected to be more qualified and competent in order to compete with males. On the other hand, they are often allowed to advance only if they subscribe to the status quo and do not question the existence of sex discrimination. Both a male approach of competition (to be twice as good in order to be valued) and a female approach of submissiveness (to subscribe to the status quo) seem to be expected of female faculty members. Morrison, et al. (1992) suggested that these conflicting demands can be combined in a Venn diagram to illustrate the narrow band of acceptable behaviors available to women wanting to break through the glass ceiling. This diagram is reproduced in Appendix B. How these expectations for the behavior of females in the workplace relate to the literature on sex-role stereotypes might offer some insight into the operation of the glass ceiling effect.

Sex-Role Stereotypes

According to Etaugh and Riley (1983), "considerable evidence exists that women are often judged to be less competent than men with equal qualifications and accomplishments" (p. 943). Assuming that all else is equal, women are being evaluated as less competent than men. This would seem to support the notion that sex-role stereotypes are contributing to the devaluation of women in work settings and therefore to the glass ceiling effect.

One reason for the devaluation of women in such settings may be related to the narrow band of acceptable behaviors posited by Morrison et al. (1992). She indicated
that only certain masculine and certain feminine behaviors were acceptable for women in the workplace. However, she did not define "masculine" or "feminine." Drawing from the work of Bem, Borders and Fong (1984) defined masculine as reflecting "an instrumental (cognitive, active) role that includes a goal-directed, task-oriented, and independent perspective" (p. 59). Feminine sex-roles, on the other hand, are defined as "an expressive (nurturant, receptive) role characterized by an affective, supportive, other-focused perspective" (Borders & Fong, 1984, p. 59). These authors reviewed the literature and reported that androgynous individuals, those manifesting a large number of masculine and feminine characteristics, are viewed as more competent than individuals manifesting primarily only one sex-role orientation.

Other research conflicts with these findings, suggesting that masculine behaviors are consistently viewed as representing better adjusted and healthier individuals. Much of the research that has been done on sex-role stereotypes and the evaluation of women has been based on the classic study by Broverman, Broverman, Clarkson, Rosenkrantz and Vogel (1970). They found that clinicians within the field of psychology base their evaluations of people's psychological adjustment on sex-role stereotypes. Specific findings indicated that clinicians perceive the characteristics of a healthy person as matching the characteristics of a healthy male (stereotypically masculine characteristics), but as not matching the characteristics of a healthy female.

A replication of the study by these same authors supported the original findings (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972). The implications of the study are enormously important. To begin with, this supports the hypothesis that a
double standard of health exists. This double standard may shed insight about the double standard that seems to exist for women in the workplace.

On the face of it, the finding that clinicians tend to ascribe the male-valued, competency cluster traits more often to healthy men than to healthy women may seem trite. However, a consideration of the content of these items reveals a powerful negative assessment of women. In effect, clinicians are suggesting that healthy women differ from healthy men by being more submissive, less independent, less adventurous, less objective, more easily influenced, less aggressive, less competitive, more excitable in minor crises, more emotional, more conceited about their appearance and having their feelings more easily hurt." (Broverman, et al., 1972, p. 70)

This would seem to place women in an awkward position when attempting to be seen as well-adjusted, both by clinicians and in the workplace. Broverman, et al. (1972) addressed this position when they described the implications for women. "If women adopt the behavior specified as desirable for adults, they risk censure for their failure to be appropriately feminine; but if they adopt the behaviors that are designated as feminine, they are necessarily deficient with respect to the general standards of behavior" (p. 75).

Perhaps as a result of heightened awareness of the impact of such stereotypes, more recent replications of the study have failed to consistently find a statistically significant bias against females by clinicians (Page, 1987; Phillips, 1985; Phillips & Gilroy, 1985; Poole & Tapley, 1988; Silvern & Ryan, 1983). As Poole and Tapley (1988) indicated, "clinical psychologists no longer hold vastly different expectations for males and females" (p. 270). Phillips and Gilroy (1985) offered other possible explanations, however. They asserted that, "perhaps clinicians have simply become more aware of the need to appear egalitarian and have not changed their 'working' bias" (p. 191). Other explanations include the idea that the transparency of analogue studies...
allows for more socially desirable responses to the study and the possibility that the women's movement has indeed resulted in some real change by clinicians.

Interestingly, though, if change has occurred, it seems to be limited to judgments of mental health. Poole and Tapley (1988) concluded that "the most striking finding is that clinicians expect both men and women to adjust their behavior to the environment, with more traditionally masculine behavior associated with the work environment and traditionally feminine behavior expected in the home environment" (p. 270). This finding is consistent with the literature on females working in traditionally male occupations. Etaugh and Riley (1983), for example, found that individuals, especially females, are devalued when perceived as working in gender-atypical occupations.

It is therefore important to examine the impact of college or university teaching being considered a masculine occupation on the advancement of women. There can be no doubt that female professors advance at a slower rate and to a lower level than do male professors. To understand this fact, Sandler and Hall (1986) suggested the necessity of considering the impact of a person's sex on the evaluations of his or her competence: "Women's roles in academe, as elsewhere, are often not only stereotyped but women are also devalued. Numerous studies—many in academic settings—demonstrate how the sex of a person influences perception and evaluation of his or her behavior and achievements" (p. 6). Although such perceptions and evaluations of a professor's competence may be held by a variety of individuals at a university, the bulk of the research has focused on student evaluations of professors. This, combined with the fact that "student ratings were most frequently mentioned as the method of teacher
evaluation used by almost 600 colleges and universities" (Gustad, 1961; cited by Elmore & La Pointe, 1974, p. 386), provides the rationale for examining the research findings on student evaluations of professors. Before doing so, however, it is necessary to understand the policies underlying evaluation within higher education.

Faculty Evaluations in Higher Education

Criteria for Tenure

A review of the literature reveals that teaching effectiveness represents only one of several criteria considered in faculty evaluations. Most consistently, three criteria are utilized by universities in the evaluation of professors: (1) research, (2) teaching, and (3) service (Centra, 1979, 1993). Other researchers have developed more comprehensive lists of criteria which should be used in faculty evaluations (Coufal & Hines, 1976; Roueche & Baker, 1987), but a careful examination of these various taxonomies will reveal that research, teaching and service are consistently emphasized.

Universities therefore tend to be consistent in their use of these three criteria in faculty evaluations. Professors are expected to demonstrate competence as researchers (most often by successful publication and/or grant-writing efforts), to show evidence of effective teaching, and to provide significant service to the university (often through committee work) and to the community.

How universities weigh the relative importance of these criteria, however, is less consistent (Coufal & Hines, 1976; Martin, 1984; Theodore, 1986). Coufal and Hines
(1976) observed "while authors give different terms and relative weights to each criteria, there is no doubt that there is strong agreement on the criteria that should be used in evaluating faculty" (p. 9). Theodore (1986) also pointed to the lack of clarity at many universities regarding the relative importance of these criteria. Reflecting this lack of clarity are statements by researchers in the field. Coufal and Hines (1976), for instance, suggested that teaching effectiveness is generally recognized as the most crucial element of faculty evaluations whereas Lin, McKeachie and Tucker (1984) asserted that "most colleges and universities attach great importance to teaching ability in criteria for promotion, yet faculty members everywhere believe that the formal criteria claim more weight for teaching than is actually accorded it" (p. 583). Finally, Martin (1984) suggested that this reflects a systematic bias which serves to give tenure to men more often than to women: "the reward structure places greater value on traditional male activities [research and administration] than on female activities [teaching and committee work]" (p. 483).

The policy statement of the department in which this research was conducted specifies the criteria which are to be used in faculty evaluations for the purposes of tenure and promotion decisions. These may be seen in Figure 1. Teaching effectiveness, therefore, is officially accorded the highest level of importance in the department in which this research was conducted. Although other researchers have suggested that the actual importance of teaching evaluations may be less than what is represented in this policy statement, to consider the implications of such a possibility in the design of this study is beyond the scope of this dissertation.
Professional Competence (50%)
- Teaching Performance
- Advising Performance
- Curriculum Contributions and Teaching Innovations
- Continuing Professional Renewal and Educational Attainments
Other Evidence of Professional Competence

Professional Recognition (30%)
- Professional Leadership and Creative Activity
- Publications, Professional Presentations, and Research
- Other Evidence of Professional Recognition

Professional Service (20%)
- Departmental committee, projects, or assigned duty
- College committee, projects, or assigned duty
- University committee, projects, or assigned duty
- AAUP Chapter committee, projects, or assigned duty
- Community Service

Figure 1. Criteria Used in Faculty Evaluations for Tenure and Promotion Decisions.

The next section will therefore address the literature regarding the purposes and sources of teaching evaluations.

**Evaluation of Teaching Effectiveness**

There exist today two very different purposes of teacher evaluation within higher education (Centra, 1979, 1993; Coufal & Hines, 1976; Cross, 1988; Root, 1987). The conceptualization of these differing purposes has been based on research done in the field of program evaluation. Within this field, "Scriven (1967) first distinguished between the
formative and summative roles of evaluation" (Worthen & Sanders, 1987, p. 34).

The formative use of teacher evaluations is designed to help faculty members improve and further enhance their teaching skills. Typically, these evaluations are informal, are often conducted by the actual professor, and are administered during a semester (as opposed to at the end). Formative evaluations of teaching effectiveness are designed to assist the instructor in determining what is working well and not-so-well, in pinpointing needed changes or improvements, and in guiding changes to make one's teaching more effective.

However, simply administering an evaluation during a semester (with time remaining for improvement) does not serve a formative purpose. Only those evaluations which are effectively used to improve instruction may be described as serving such a purpose. As Centra (1993) indicated, evaluations may serve a formative purpose only if the following four conditions are met:

First, teachers must learn something new from them. Second, they must value the new information. Third, they must understand how to make improvements. And, finally, teachers must be motivated to make the improvements, either intrinsically or extrinsically. (p. 81)

The second purpose of faculty evaluations presently utilized in universities is summative. Summative evaluations are conducted for the express purpose of collecting data which will be used in tenure and promotion decisions (Centra, 1993). In collecting information about a faculty member's teaching effectiveness for summative purposes, universities have relied on several sources of information. These sources include student evaluations, self-evaluations, peer (or colleague) evaluations, and alumni evaluations
Self-evaluations involve the professor rating himself or herself. The use of self-evaluations has traditionally served more formative purposes, with the instructor identifying areas of strength and areas in which growth or improvement is needed. This information is then used to set goals. Because this study is interested in evaluations used for summative purposes, self-evaluations will not be considered. Furthermore, this study will limit its consideration of summative evaluations to an examination of student evaluations.

Peer evaluations involve professors rating the teaching effectiveness of other professors. However, researchers have suggested that peer evaluations tend to be less standardized and therefore rather idiosyncratic in terms of their influence on tenure and promotion decisions (Centra, 1975; Marsh, 1984). One reason for this is that peer evaluations tend to be based on a limited number of observations, to involve informal reports of perceived effectiveness, and to often be based on informal comments by students. Therefore, although such informal evaluations may actually have considerable influence on the tenure and promotion decisions made by universities, the difficulties involved in systematically collecting and analyzing the data are immense. This researcher therefore chose not to target this variable.

Alumni evaluations involve graduates rating the teaching effectiveness of professors retrospectively. Interestingly, attention to alumni evaluations was prompted by research into the reliability of student evaluations. Because one criticism of student evaluations is often that students are immature and lack the foresight necessary to
appreciate the quality of education they received, much research was done on the stability of student evaluations over time. This research will be discussed in the upcoming section on reliability.

Centra (1979, 1993) has indicated that student evaluations are used by most institutions as their primary summative measure of teaching effectiveness. Cashin (1989) and Marsh, Overall, and Kesler (1979) concur that student evaluations are used by most institutions as their primary summative measure of teaching effectiveness. The next section of this literature review will therefore focus on the use of student evaluations of teaching effectiveness. This section will specifically address the history and development of student evaluations, the measurement of student evaluation, and research on the accuracy of student evaluations. The following section will then examine the findings regarding the influence of sex and gender-role orientation on student evaluations which has been documented in the literature.

Student Evaluations of Teaching Effectiveness

History and Development of Student Evaluations

The use of student evaluations in higher education has a long and varied history. Some have suggested that student evaluations originated during medieval times in Europe (Rashdall, 1936; cited by Centra, 1993). These evaluations took two forms: (1) the monitoring of the instructor's adherence to a schedule of lecture topics, and (2) the practice of having students pay admission directly to the professor rather than to the
university. Students who evaluated a professor's lectures as extremely poor would be unlikely to attend and therefore would result in a direct "pay cut" to those professors who were negatively evaluated (Centra, 1993).

Ory (1990) suggested that student evaluations within the United States first became popular in the 1920's. At this time, according to Ory, the evaluations were initiated by various student groups for the purpose of assisting students with the selection of courses and instructors. Faculty participation at this time was voluntary. Centra (1993) noted that research on student evaluations also began during this period and concurred with Marsh (1987) in his identification of Herman Remmers as the "Father of Student Evaluation Research" (Centra, 1993, p. 49). Much of this research was conducted at Purdue University and investigated the reliability of student responses to the Purdue Rating Form. This form, which was published in 1927, is recognized as the "first student evaluation form" (Centra, 1993, p. 49).

This early period in the use of student evaluations was therefore marked by student initiation, by voluntary faculty participation, and by early research on reliability. Centra (1993) reported that use of these student evaluations was limited during this early period which spanned over three decades (1920's to 1960).

The 1960's, however, brought about dramatic changes in the use of student evaluations and is therefore recognized by both Centra (1993) and Ory (1990) as the second period in the "modern era of student evaluations" (Centra, 1993, p. 490). Students on campuses nationwide were protesting in the 1960's not only in reaction to national issues such as the Vietnam war but also in reaction to university policies. Centra
(1993) explained, "Increasingly, students saw themselves as consumers. They demanded a voice in governance; they wanted to improve the education they were receiving. Evaluating their courses and teachers was one way to make their voices heard" (p. 50). The use of student evaluations therefore became more prevalent during the 1960's. However, few universities considered the results of these evaluations when making tenure and promotion decisions.

It was not until the 1970's, which Centra (1993) recognized as the third period of student evaluations, that universities began to use student evaluations on a widespread basis. Ory (1990) identified several factors which contributed to the adoption of student evaluations by university administrations: "In the 1970's, increased costs of higher education and mounting financial problems added to students' cries for acceptability and led campus administrators to consider using systematically collected student ratings of instruction in the decision-making process" (p. 64).

At the same time, however, faculty members remained unconvinced that student ratings were both reliable and valid measures of an instructor's effectiveness. As a result, the 1970's were also a time of intense research into these issues. Much of this research was conducted by "campus instructional services and evaluation offices" (Ory, 1990, p. 64). Researchers such as Centra, who was instrumental in developing the Student Instructional Report (SIR), and Marsh, who developed the Students' Evaluations of Educational Quality (SEEQ), dominated the field of research on student evaluations at that time. As Centra (1993) observed, the effect of such research was that "the generally favorable findings helped support the use of the evaluation for tenure and promotion
decisions" (p. 50).

By the 1980's, therefore, most universities were systematically using student evaluations for summative purposes. Seldin (1984) reported that approximately 70 percent of universities were collecting student evaluations at that time, and Ory and Parker (1989) indicated that this percentage has increased to 100 percent by the end of the decade. The following section will describe characteristics of the various student evaluation instruments being utilized.

Measurement of Student Evaluations

As just noted, there exists an apparently widespread reliance by universities on student evaluations as a measure of teaching effectiveness. However, although student evaluations tend to be administered systematically (Cashin, 1989), they are not to be considered a uniform measure. In fact, there are a variety of instruments with which universities measure student evaluations of teaching effectiveness. In addition, each instrument tends to target a variety of separate measures. Before addressing the various instruments available for student evaluations, this section will next address the types of measures most often contained within each instrument.

Global Versus Specific Measures

Student evaluations are frequently designed to elicit information from the respondents about the quality of the instructor in a number of respects. Some of the categories of information are designed to target specific dimensions of teaching...
effectiveness. For instance, Centra (1993) listed the following factors:

1. Organization, planning, or structure
2. Teacher student interaction or rapport...
3. Clarity, communication skill
4. Work load, course difficulty
5. Grading and examinations, assignments
6. Student learning, student self-ratings of accomplishments or progress. (p. 57)

Marsh (1984, 1991a) advocated the use of an instrument which measures the following nine separate dimensions of teaching effectiveness:

1. Learning/Value...
2. Instructor Enthusiasm...
3. Organization...
4. Group Interaction...
5. Individual Rapport...
6. Breadth of Coverage...
7. Examinations/Grading...
8. Assignments... [and]

Similarly, Feldman (1988) identified the following twenty-two "instructional dimensions" of effective teaching in his review of research in the area:

1. Teacher's Stimulation of Interest in the Course and Its Subject Matter...
2. Teacher's Enthusiasm (for Subject or for Teaching)...
3. Teacher's Knowledge of Subject...
4. Teacher's Intellectual Expressiveness (and Intelligence)...
5. Teacher's Preparation; Organization of the Course...
6. Clarity and Understandableness...
7. Teacher's Elocutionary Skills...
8. Teacher's Sensitivity to, and Concern with, Class Level and Progress...
9. Clarity of Course Objectives and Requirements...
10. Nature and Value of the Course Material (Including Its Usefulness and Relevance)...
11. Nature and Usefulness of Supplementary Materials and Teaching Aids...
12. Perceived Outcome or Impact of Instruction...
13. Instructor's Fairness; Impartiality of Evaluation of Students; Quality of Examinations...
Thus, student evaluation instruments tend to be designed in such a way as to measure a number of specific dimensions thought to be correlated with effective teaching. These specific dimensions tend to be especially helpful in the formative evaluation process because they allow teachers to better understand what students liked and disliked about their teaching style.

Evaluation forms also tend to include measures of global assessments of professor competence. These items tend to refer to overall effectiveness of the teacher or the overall quality of the course. Examples of global items on the Instructional Development and Effectiveness Assessment (IDEA) assessment instrument (Hoyt & Cashin, 1977) are "Overall, I rate this INSTRUCTOR an effective teacher" and "As a result of taking this course, I have more positive feelings toward this field of study."

However, although global ratings tend to correlate highly with a number of specific factors (Centra, 1993), there exists an ongoing debate about the appropriate uses of each. While researchers generally agree that ratings of specific qualities are more useful than global ratings for formative purposes (Abrami, 1988, 1989; Cashin, 1990b;
Cashin & Downey, 1992; Centra, 1979, 1993; Marsh, 1991a, 1991b), they continue to
debate about whether specific (multidimensional) ratings or global ratings are more
useful for the purpose of summative evaluations.

As Cashin and Downey (1991) observed, "one of the continuing debates
concerning the use of student ratings of teaching is the debate revolving around what
kind of measures should be used for summative evaluation of faculty, in making
personnel decisions for retention, promotions, tenure, or salary increases, and of course,
to assess their effectiveness" (p. 563). On one end of the continuum of opinions are
Abrami (1988), Abrami and d'Apollonia (1991), Cashin (1990b), Cashin and Downey
(1992), and Centra (1979, 1993). These researchers have argued that only global items
should be used for summative decisions regarding tenure and promotion. The rationale
behind advocating only the use of global ratings for summative purposes was summarized
by Abrami and d'Apollonia (1990). First, they questioned the content validity of specific
items when "used across a wide variety of courses, instructors, students, and settings"
(Abrami & d'Apollonia, 1990, p. 98). For example, specific items inquiring about
whether the instructor encourage all students to participate in class discussions has an
obvious lack of relevance in a large lecture course and it would therefore be
inappropriate to find the instructor to be less effective due to a failure to encourage the
active participation of well over one hundred students in class discussions. Cashin and
Downey (1992) also discussed this difficulty with specific measures of effective teaching
in relation to the development of the Instructional Development and Effectiveness
Assessment (IDEA) evaluation form: "One critic rather forcefully suggested that many
of the items could reflect both bad teaching and good teaching, observing, for example, that 'well organized garbage still smells' (Hoyt, 1973a, p. 153)" (Cashin & Downey, 1992, p. 564).

A second argument against the use of specific ratings in summative decisions and for the exclusive reliance on global measures is based on correlational research. Research findings have tended to yield higher correlations between student learning and global measures than between student learning and specific measures (Cohen, 1981; cited in Abrami & d'Apollonia, 1990; Cashin & Downey, 1992).

Abrami and d'Apollonia (1990) pointed to two additional reasons for preferring global assessments to specific measures for summative decisions. First, "much less is known about the generalizability of specific rating factors than about global ratings" (p. 98). Secondly,

one cannot expect untrained administrators or nonexperts in evaluation, attempting to arrive at a single decision about the quality of an instructor's teaching, to properly weigh the information provided by factor scores (Franklin and Theall, 1989)....It would be particularly disappointing to learn that a faculty member was denied tenure because of low student ratings on Difficulty when such ratings are almost totally uncorrelated with students' learning. (Abrami & d'Apollonia, 1990, p. 99)

Other researchers disagree with the assertion that only global measures should be used for summative decisions. On the other end of this continuum are researchers who assert that only specific ratings should be used, both for summative and formative purposes. Marsh (1991b) summarized this position and associated this stance primarily with Frey:

At the opposite extreme Frey (1973, 1974, 1978; Frey & Flay, 1978; Frey, et al.,
argued strongly that only specific dimensions should be considered, and he excluded global ratings from his Endeavor instrument. His subsequent research on two higher order dimensions (Frey, 1978) led him to conclude "that personnel decisions should not be made on a single global evaluation measure (Frey & Flay, 1978, p. 25). Frey's main arguments were that (a) global items are too much influenced by variables that are not associated with effective teaching; (b) global ratings are unduly influenced by SETE components that are minimally related to student achievement, which Frey argues is the most important criterion of effective teaching; and (c) it is better to focus on components that are maximally related to a particular criterion than to rely on global items. (p. 419)

Marsh (1991b) is alone, however, in his identification of Frey as being at the other end of the continuum. Abrami and d'Apollonia (1990), Cashin (1990b), and Cashin and Downey (1992) identify Marsh as representing the other extreme. As Cashin and Downey (1992) stated,

During the past decade, probably no one has championed the case for the multidimensionality of student ratings more persuasively than Marsh (1984, 1987, 1991a, 1991b)....[He] contends that because students' ratings are multidimensional (e.g., an instructor may be well organized but lack enthusiasm), multiple ratings should be used...and has specifically raised the issue of using student ratings for summative versus formative evaluation. (p. 563)

A debate between Abrami and d'Apollonia and Marsh in the Journal of Educational Psychology provides additional evidence that, although Marsh perceives himself as representing a "middle ground between the extreme positions proposed by Abrami and d'Apollonia...and Frey..., recommending the use of both specific dimensions and global ratings" (Marsh, 1991b, p. 419), Marsh has been identified by the proponents of using only global evaluations for summative decisions as the primary opponent of their opinion.

Nonetheless, a careful reading of Marsh will attest to the fact that he does, regardless of being perceived as representing the opposite extreme, indeed take a moderate position. Marsh and Bailey (1993) stated, "Feedback from global ratings may
provide a reasonable indicator of overall effectiveness from the perspective of students, but they provide little or no diagnostic value about what specific areas of teaching are in need of improvement" (pp. 11-12). They continue by summarizing Marsh's (1991b) response to Abrami and d'Apollonia in *the Journal of Educational Psychology:

Marsh [15, p. 419] noted that "because this is valuable information, SEEQ contains global ratings and I support their use." However, consistent with Frey's perspective, Marsh [15, p. 419] also argued "that there are many criteria of teaching effectiveness, that each criterion will be differentially related to different SETE dimensions, and that any criterion can be inferred more accurately with an appropriately weighted set of specific dimensions that with a global rating item."

The use of a weighted average of specific dimensions may represent a compromise that is consistent with recommendations by Abrami, by Frey, and by Marsh. (p. 12-13)

Although this debate may eventually be resolved, it is necessary at this point to be aware of its existence and the implications of each perspective. One such implication is related to the selection and interpretation of standardized student evaluation instruments. Although some universities utilize student evaluation instruments which are developed on-site, a number of published instruments are available for purchase and are widely used by universities for the purpose of measuring student evaluations of professor teaching competence. These instruments tend to be associated with researchers in the area of educational assessment described above and tend to reflect the positions of their developers. The next section will therefore address several widely used student evaluation instruments.

**Student Evaluation Instruments**

This section will begin with a description of the Instructional Development and
Effectiveness Assessment (IDEA) instrument and the Student Instructional Report (SIR). According to Centra (1979; 1993) and Cashin (1990a), these two instruments are "probably the two most widely used student rating systems in North America" (Cashin, 1990a, p. 114). Next, the Student's Evaluation of Educational Quality (SEEQ) will be described in similar depth. This section will then turn to a brief description of a number of other instruments designed to measure teaching effectiveness from the student perspective. Both an exhaustive review of all available student evaluation of teaching effectiveness (SETE) instruments and a detailed analysis of the instruments described in this section are beyond the scope of this review of the literature. More detailed information about the strengths and weaknesses of the available instruments may be found in Centra (1979, 1993) and in the manuals for each individual instrument.

**Instructional Development and Effectiveness Assessment (IDEA)**

The Instructional Development and Effectiveness Assessment (IDEA) System is a widely utilized and well studied instrument (Cashin, 1990b; Centra, 1993). This instrument was developed by the Center for Faculty Evaluation and Development at Kansas State University, was first published in 1977, and was most recently revised in 1988. It is a 46 item self-report inventory which inquires about the students' reactions to the instructor and to the course and which also elicits information regarding the students' perceptions of their progress with regard to a wide range of instructional goals.

The developers of the IDEA System (Hoyt & Cashin, 1977) have perhaps been the most enthusiastic supporters of the use of student learning as the criterion for
effective teaching. Their argument has been that the logical result of effective teaching is student learning. Although Marsh (1984) identified student learning as "the most widely accepted criterion of effective teaching," (p. 720) the IDEA System is the only widely used student rating instrument which employs student learning as the major criterion for teaching effectiveness. The IDEA System is therefore unique as a student evaluation instrument because of its use of student learning as the primary criterion for professor effectiveness.

In addition to providing a global measure of teaching effectiveness based on student self-reported progress with respect to learning objectives, the IDEA form also provides an global rating of overall teacher effectiveness and a global rating of overall course quality. Finally, the instrument contains a number of items designed to elicit student ratings about specific teacher behaviors. Consistent with the preceding discussion, Cashin (1990b) recommends that only the global measures be used for summative purposes and suggests that the specific measures be used diagnostically for formative purposes.

Three strengths of the IDEA rating system are its widespread use in North America, allowing for normative data on a national basis, the fact that it is a widely researched instrument, and its inclusion of student self-reported achievement as an indicator of teacher effectiveness.

**Student Instructional Report (SIR)**

The Student Instructional Report (SIR) is also a widely used and well studied
instrument (Cashin, 1990b; Centra, 1993). It was developed by the Educational Testing Service (Centra, 1973) and was first published in 1971. It was most recently revised in 1989. The SIR rating system is a self-report inventory which consists of "thirty-nine questions, plus space for responses to ten additional questions that may be inserted locally" (Centra, 1993, p. 188).

The SIR is designed to elicit student opinions regarding specific characteristics and behaviors of the teacher as well as to ascertain the students overall opinion regarding a variety of global qualities. Centra (1979, 1993) recommends that the global items be used for summative decisions and that the more specific items be used for diagnostic and formative purpose.

Two strengths of the SIR include its widespread use in North America, allowing for normative data on a national basis and the fact that it is a widely researched instrument.

**Student's Evaluation of Educational Quality (SEEQ)**

The Student's Evaluation of Educational Quality (SEEQ) is a well studied instrument developed in Australia by Marsh and first published in 1976 by Marsh. It was most recently revised in 1991 and consists of thirty-five items designed to measure the following nine evaluation factors:

1. Learning/Value...
2. Instructor Enthusiasm...
3. Organization...
4. Group Interaction...
5. Individual Rapport...
In addition to assessing these nine specific areas of teaching effectiveness, the SEEQ also includes three items designed to assess the student's perception of the overall, global quality of the class and the teacher. However, consistent with the preceding discussion regarding the use of specific and global measures for summative decisions, Marsh does not advocate the sole use of global items for summative decisions (Marsh, 1984, 1987, 1991a, 1991b).

Two strengths of the SEEQ are the fact that it is well researched and that it is available free of charge. A weakness is related to recommendations for summative evaluations. Because Marsh recommends a complex approach to the making of summative decisions, a level of administrator expertise in evaluation is generally advisable.

To supplement his research on the SEEQ in Australia, Marsh has collaborated on researching the SEEQ in the United States (Marsh & Hocevar, 1991; Marsh & Overall, 1981; Marsh, Overall, & Kesler, 1979). Based on this research and the use of the SEEQ in the United States, there exist two forms of the SEEQ: "an 'Australianized' version...incorporating minor modifications to reflect Australian spelling and usage" (Centra, 1993, p. 204) and a version in standard, American English.
Endeavor Instructional Rating System (EIRS)

The Endeavor Instructional Rating System (EIRS) was developed by Frey at Northwestern University and reflects his belief in using only specific rating items with no global assessment items. It consists of 7 items and provides ratings of the teacher's organizational skills, communication skills, interpersonal skills, and the difficulty of the course. The instrument is not published and is available directly from Frey.

Instructor and Course Evaluation System (ICES)

The Instructor and Course Evaluation System (ICES) "is a computer-based system through which faculty can select items from a catalogue of more than 400 items classified by content...and specificity...(Centra, 1993, p. 181). Only the global and general concept items are normed; the specific items are not normed as they are recommended only for formative use. This instrument is available from the University of Illinois at Urbana.

Student Instructional Rating System (SIRS)

The Student Instructional Rating System (SIRS), not to be confused with the SIR published by Educational Testing Service, was developed at Michigan State University. The standard form consists of twenty items related to specific ratings and one general, global affect item (asking students to rate their "general enjoyment of the course"). The instrument was copyrighted in 1982 and is available for use by other universities.
Arreola and Heinrich (1977) reported that Florida State University had also adopted the SIRS for use on its campus.

**Instructional Assessment System (IAS)**

The Instructional Assessment System (IAS) actually consists of nine separate student evaluation forms. The form to be utilized depends upon the structure and type of course being evaluated. Centra (1993) reported that forms exist for "large lecture, small lecture-discussion, seminar, problem-solving course, skill acquisition course, quiz section, homework section, lab section, and clinical section" courses (pp. 179-180). The instrument yields both global and specific ratings and is available from the Educational Assessment Center at the University of Washington.

**Research on the Accuracy of Student Evaluations**

Despite the widespread availability and use of standardized student evaluation instruments for both summative and formative purposes, faculty members have historically objected to their use in summative decisions. As Marsh, Overall, and Kesler (1979) observed, "While few faculty argue strongly against the usefulness of [student] ratings in providing feedback about instructional effectiveness to the faculty themselves, many continue to challenge the use of such ratings in personnel decisions" (p. 149). Frequently, these objections have been based on charges regarding insufficient reliability or validity of the instruments used. Concerns about the validity of student evaluation instruments have been voiced in allegations of bias. In fact, "the most common criticism
of student evaluations is that they are biased by variables unrelated to teaching effectiveness" (Marsh, Overall, & Kesler, 1979, p. 149).

In large part because of the passionate objections to the use of student evaluations of teaching effectiveness for summative decisions, there has been a great deal of research conducted regarding the accuracy of student evaluations. This section will therefore specifically address studies regarding the reliability and validity of student evaluations in general. The first subsection will examine reliability research; the second subsection will address validation research; and the third subsection will address research regarding potential sources of bias.

**Reliability of Student Evaluations**

Research on the reliability of student evaluation instruments began in the 1920's and was originally conducted by Remmers. This research is most often specific to a single instrument and focuses on three areas of reliability: (1) consistency or interrater reliability, (2) stability, and (3) generalizability. Little consideration is given to reliability estimates based on measures of internal consistency. As Marsh (1984) pointed out, internal consistency estimates of reliability are usually inflated because "they ignore the substantial portion of error due to lack of agreement among different students" (p. 716).

It is this agreement (or lack thereof) among students, the interrater reliability or consistency, that has been of the most interest to researchers studying the reliability of student evaluations. Because this research tends to be specific to individual instruments, the reader is advised to consult the manual for the particular student evaluation
instrument to be used. However, research done on various instruments has yielded similar results.

As a general finding within this research, interrater reliability of student evaluations appears to vary not only with the instrument being used but also with the class size. For example, Marsh (1984) summarized the interrater reliability of the SEEQ as being .20 when referring to the degree of agreement between any two students in a class (randomly selected) and as ranging from .60 for a class size of 5 to .95 for a class size of 50 when class-average responses are considered. Similar interrater reliability figures are cited by Cashin (1988) in reference to the IDEA system: "10 raters = .69; 20 raters = .81; 40 raters = .89" (p. 1). These data indicate a generally acceptable degree of consistency given class sizes of at least 15.

The second aspect of reliability which is of interest when considering student evaluations is stability. This refers to the stability of ratings over time and can be studied cross-sectionally or longitudinally. Cross-sectional studies compare "the retrospective ratings of former students and those of currently enrolled students" (Marsh, 1984, p. 717). Longitudinal studies compare the same students' end-of-course evaluations with their evaluation of the same professor/course at least one year later.

In a cross-sectional study, Centra (1974) found "substantial agreement between current students and alumni (of five years) regarding who have been effective or ineffective teachers" (p. 321). Feldman (1989) conducted a major review and synthesis of research into the reliability of student evaluations and found substantial similarity between ratings by current students and ratings by former students.
In a longitudinal study by Overall and Marsh (1980), the stability of ratings (i.e., correlation between ratings and retrospective ratings) was .83. They concluded that "these findings demonstrate that students' evaluations collected at the end of a course are remarkably similar to the retrospective ratings provided by the same students several years later" (Overall & Marsh, 1980, p. 321). Marsh and Overall (1981) also described these findings as demonstrating a correlation substantial enough to demonstrate reliability in the form of stability. Aleamoni (1987) reported similar results supporting the reliability of student evaluations based both on consistency and on stability.

The third aspect of reliability is the generalizability of the results. Specifically, this issue addresses the level of confidence one can have that student evaluations are reflective of an instructor's effectiveness rather than an artifact of a particular course. Marsh (1981) conducted a comprehensive study of the generalizability of student ratings. In this study, he used student ratings from 1,364 classes and arranged the data into four groups: (1) ratings of the same teacher and same course, (2) ratings of the same teacher but different courses, (3) ratings of different teachers teaching the same course, and (4) ratings of different teachers teaching different courses. Marsh then examined the correlation of student evaluations of each of these four groups. Table 1 summarizes the results. Marsh (1984) explained his interpretation of these results: "Based on a path analysis of these results, I argued that the effect of the teacher on student ratings of teaching effectiveness is much larger than the effect of the course being taught" (p. 718). Marsh and Overall (1981) echoed this conclusion. Cashin (1988) concurred, stating that "Marsh's results are reasonably comparable to other generalizability studies.
Table 1

Correlation of Student Ratings: Teacher vs. Course Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Same Course</th>
<th>Different Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same Teacher</td>
<td>.707</td>
<td>.523</td>
</tr>
<tr>
<td>Different Teacher</td>
<td>.140</td>
<td>.061</td>
</tr>
</tbody>
</table>

(Bausell, Schwartz, & Purohit, 1975; Gilmore, Kane, & Naccarato, 1878; and Hogan, 1973)" (p. 2). More recently, Barnes and Barnes (1993) argued for the generalizability of student evaluations but specified that this generalizability is stronger when the evaluation is of the instructor (as opposed to the course). They concluded that "...student evaluation data appear to provide a reasonable basis for making decisions about instructors when generalizing across courses and students..." (p. 135).

The research on the reliability of student evaluations has therefore supported their consistency, stability, and generalizability. This section turns now to the research on the validity of student evaluations.

Validity of Student Evaluations

Validity refers to the extent to which student evaluations actually measure what they are intended to measure (instructor effectiveness). Validity, however, is especially difficult to research because researchers concede that there is no universally accepted criteria for what constitutes effective teaching (Aleamoni, 1987; Cashin, 1988; Feldman,
Research intended to measure the validity of student evaluations has therefore tended to compare these evaluations to a measure of student learning or to compare them to other evaluations of teacher effectiveness which are assumed to be valid. These other sources have included instructor self-evaluation, peer/colleague evaluations, administrator evaluations, and alumni evaluations.

**Correlations With Student Learning**

The developers of the IDEA System (Hoyt & Cashin, 1977) have perhaps been the most enthusiastic supporters of the use of student learning as the criterion for effective teaching. Their argument has been that the logical result of effective teaching is student learning. Therefore, the teacher whose students learn more can be described as more effective.

Although Marsh (1984) identified student learning as "the most widely accepted criterion of effective teaching," (p.720) only one widely used student rating instrument employs student learning as the major criterion for teaching effectiveness. The IDEA system is unique as a student evaluation instrument because of its use of student learning as the primary criterion for professor effectiveness and was chosen for this feature.

The IDEA System relies on student self-report of progress with respect to 10 objectives in order to measure student progress. Although Cashin and Downey (1992) cautioned that the IDEA instrument "is more a measure of student learning—actually students' perceived learning—than of teaching effectiveness per se" (p. 568), researchers have found evidence to support the validity of this approach. Specifically, Ohara and
Purcell (1980) conducted research to determine whether students' self-reports of achievement could be considered a valid measure of teaching effectiveness. They found statistically significant correlations between student self-reports of achievement and ratings of teaching effectiveness, "supporting the thesis that self-reported achievement may be substituted for actual achievement" (Ohara & Purcell, 1980, p. 1).

Other researchers have utilized actual student grades as a measure of student progress. Cohen (1981; cited by Cashin, 1988) reviewed a number of studies using student grades and found that student grades correlated only .47 with their self-reported progress, .47 with overall ratings of course effectiveness, .44 with overall ratings of instructor effectiveness. Although these validity coefficients are substantially lower than the reliability coefficients cited earlier, Cashin (1988) asserted that validity coefficients between .20 and .49 are "practically useful... especially when studying complex phenomenon, such as student learning" (p. 2). He would therefore consider these coefficients as supportive of the validity of student evaluations.

Cruse (1987), on the other hand, also focused on Cohen's (1981) findings but disagrees with Cashin's conclusion. Among his criticisms of student evaluations which call into question their validity is the fact that "the correlation between overall instructor ratings and student achievement can be .38 if the ratings are made before the students know their final grade but .85 if the ratings are made after final grading (Cohen, 1981)" (p. 729). Cruse (1987) suggests that grade expectancies, along with other factors, bias student ratings.
Correlations With Other Ratings of Teaching Effectiveness

Because faculty members such as Cruse (1987) have historically challenged the validity of student ratings, another approach to validity research has been to measure the correlation between student evaluations of an instructor and other evaluations of teaching effectiveness which are assumed to be valid. These individuals have included instructor self-evaluations, evaluations by peers/colleagues, evaluations by administrators, and evaluations by alumni.

Surprisingly, few studies have actually employed the approach of comparing student evaluations with instructor self-evaluations. The correlations between self-ratings and student ratings has ranged from .20 to .65. Marsh (1984) concluded from his review of the 10 studies using this approach that these correlations suggest that "students' evaluations show significant agreement with instructor self-evaluations [and provide] a demonstration of their validity that is acceptable to most researchers" (p. 723). In his own study, Marsh (1982) found a median correlation of .45 between student evaluations and instructor self-evaluations and concluded that "these findings demonstrate student-instructor agreement on evaluations of teaching effectiveness..." (p. 264). Other researchers who have reached similar conclusions include Overall and Marsh (1982), Marsh, Overall, and Kesler (1979), and Feldman (1988). In Feldman's (1988) study, students and teachers were asked to identify "instructional characteristics they considered important to good teaching and effective instruction" (p. 291). The average correlation of .71 was interpreted by Feldman (1988) as demonstrating "generally
similar" views of effective college teaching.

Despite this general agreement, however, not all researchers agree. In his major review and synthesis of "research comparing the [actual] overall ratings of college teachers' effectiveness made by current and former students,... and the teachers themselves," Feldman (1989) concluded that "teachers' self-ratings and current student ratings are, at best, moderately similar" (p. 137). Centra (1973), whose study was among those reviewed by Marsh (1984) and yielded a .21 correlation, also concluded otherwise: "Teachers self-ratings had only a modest relationship with the ratings given by students.... There was also a tendency for teachers as a group to give themselves better ratings than their students did" (Centra, 1973, p. 286).

Another approach to determining the validity of student evaluations has been to measure their correlation with ratings by the instructor's peers or colleagues. Many of the early studies obtained peer ratings without requiring that the ratings be based on observations during a classroom visitation. Blackburn and Clark (1975) found a correlation of .62; Maslow and Zimmerman's study (1956) yielded a correlation of .69; and Murray (1972) reported a correlation of .87. Although these findings are at first glance rather impressive, researchers have questioned their value and suggested that the peer ratings in these studies were likely to have been based on conversations with students or on knowledge of an instructor's student ratings (Centra, 1975; Marsh, 1984).

More recently, validation research involving peer ratings has required the ratings to be based on classroom visitations. Another study by Centra (1975) found a correlation between student ratings and peer ratings of only .20. Feldman (1988) and
Marsh (1984) conducted reviews on the research using peer ratings with classroom visitations and both concluded that the correlation between peer ratings and student ratings is unacceptably low. Marsh (1984) stated "Peer ratings based on classroom visitation do not appear to be substantially correlated with student ratings or with any other indicator of effective teaching. Although these findings neither support nor refute the validity of student ratings, they clearly indicate that the use of peer evaluations of university teaching for personnel decisions is unwarranted" (p. 725).

Also calling into question the overall validity of peer ratings, Centra (1975) found that the correlation among peer raters was only .26. Although Root (1987) found a relatively high interrater reliability between peers and used these results to argue against the use of student evaluations, she concurred with Centra in stating that "the interrater reliability in classroom observation by peers is typically much lower than the levels reported here, low enough to suggest extreme caution in their use in summative evaluations (Centra, 1979)" (Root, 1987, p. 81).

Therefore, although peer ratings are sometimes offered as a valid measure of teaching effectiveness and used in validation research for student evaluations of teaching effectiveness, this approach has not been adequately supported by the available empirical research.

Yet another approach to determining whether student ratings actually measure what they are intended to measure (instructional effectiveness) is to measure the correlation between student ratings of an instructor at the end of a course and then following the students' graduation. The premise underlying this approach is that "follow-
up ratings allow former students to develop additional perspectives about, and to obtain emotional distance from, the person and situation being assessed," thus enabling former students to offer "more informed and mature perceptions of course and instructor effectiveness" (Overall & Marsh, 1980, p. 321). Based on this line of logic, the long-term consistency of student ratings is often offered as evidence of both the reliability and the validity of student ratings (Marsh, 1984). As indicated in the reliability section, these correlations have consistently ranged from .75 (Centra, 1974) to .83 (Overall & Marsh, 1980).

The conclusions regarding the validity of student ratings based on these four approaches to gathering data vary widely. Cashin (1988), it will be recalled, indicated that validity coefficients between .20 and .49 are high enough to be practically useful. Given this assumption and the research reviewed, it can be concluded that student evaluations are valid measures of instructor effectiveness. Although this conclusion has been offered by Cashin (1988), Marsh (1984), and Feldman (1988, 1989), some researchers have called such a conclusion into question. Dowell and Neal (1982), for example, reviewed research on the validity of student ratings of teaching and concluded that "the literature can be seen as yielding unimpressive estimates of the validity of student ratings....modest at best and quite variable" (p. 59). Abrami, d'Apollonia, and Cohen (1990) also reviewed the literature on the validity of student ratings and noted that the conclusions reached by the various researchers varied immensely:

These reviews have reached markedly different conclusions about the validity of student ratings, ranging from "strong support for the validity of student ratings as measures of teaching effectiveness" (P.A. Cohen, 1981, p. 300) to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Because of the widely discrepant opinions and research findings with regard to the validity of student evaluations, it is essential to consider the reasons for such divergent conclusions. One reason may be due to statistical issues. Because the validity coefficients Cashin (1988) cited as "practically useful" may yield statistically significant results, his conclusion may be warranted. However, it is also essential to question whether the effect size is also practically useful. In considering validity coefficients of .20 to .49, it is important to recognize that the corresponding coefficients of determination (an estimate of effect size) range from 0.04 to 0.24, respectively. This suggests that the proportion of variability in student ratings explained by variability in student progress or learning, in self-evaluations, in peer evaluations, or in alumni evaluations tends to be between .04 and .24. Although the correlations may be of statistical significance, it is also legitimate to question other sources of variability.

Another possible explanation for the discrepant opinions regarding the validity of student evaluations and for the inconsistent results with regard to sex bias is found in a statement by Machina (1987):

In general, student evaluations can be taken to report honestly student perceptions.... Perceptions are not necessarily accurate representations of the objective facts, but they nevertheless constitute, for a variety of important factors, the entirety of the student end of the teaching process. Thus, their importance in the teaching-learning interchange should be obvious. (p. 20)

Machina (1987) continued:

An intelligent use of student evaluations looks to those evaluations solely for information relating to effectiveness in reaching the students, even when the
questions on the evaluation form ask the students to rate the instructor's competence in the field, course organization, and the like. (p. 22)

When viewed in this context, validity would refer not to whether the student evaluations of teaching effectiveness measured the actual, objective quality of teaching, but rather to whether student evaluations accurately reflect the degree to which, in Machina's words, the student was reached, a criterion which is obviously subjective. This latter criterion for effective teaching once again highlights the primary difficulty in doing validity research on evaluations of teaching effectiveness: researchers concede that there is no universally accepted criteria for what constitutes effective teaching (Cashin, 1988; Feldman, 1988; Marsh, 1984).

Finally, other sources of variability may explain the discrepant conclusions regarding the validity of student evaluations. This section will now turn to a brief examination of research which has been conducted on a variety of factors which have been hypothesized to represent bias and to explain variability in student ratings which is not due to actual teaching effectiveness.

Before proceeding to the next subsection, however, it is necessary to define bias. Preece (1990) offered the following definition: "In statistics, a bias is a latent influence that disturbs an analysis. In students' evaluations of teaching effectiveness, a bias is a systematic error in student ratings" (p. 5). Centra (1993) defined bias as "a circumstance that unduly influences a teacher's ratings, although it has nothing to do with the teacher's effectiveness" (p. 65). Using both definitions, bias in student evaluations can be considered a latent circumstance which systematically influences student perceptions of
teaching effectiveness but is unrelated to actual teaching skill.

Research on Potential Bias

Research on potential biases has traditionally been conducted in three areas: (1) instructor characteristics, (2) student characteristics, and (3) course characteristics. The research in these areas has been conducted to determine whether characteristics of the student, course, or instructor which are unrelated to teaching skill may explain variability in student ratings.

Student Characteristics

There has been relatively little research conducted on the impact of student characteristics on student evaluations of teaching effectiveness. With the exception of research on the potential influence of student sex on student evaluations of teaching effectiveness, this research has tended to grow out of concerns about students not being mature enough to accurately evaluate teaching effectiveness. As such, the research has primarily addressed student sex and/or gender-role orientation; and student age, maturity, and/or class level.

With regard to the potential influence of student sex and/or gender-role orientation on student evaluations of teaching effectiveness, the research has yielded unimpressive findings. Although a number of studies have found student evaluations to be affected by an apparent interaction between the sex or gender-role orientation of the student and the instructor sex or gender-role orientation (Bray & Howard, 1980;
Feldman, 1993; Kaschak, 1981; Lueck, Endres, & Caplan, 1992; Overall & Marsh, 1982), the actual impact of these effects seems to be relatively small. Overall and Marsh (1982) summarized this research by stating that "There is some evidence that when the gender of both student and instructor is the same, higher evaluations may result on some teaching dimensions" (p. 10). Researchers finding no impact of student sex include Arden (1989), Cashin (1990b), Centra (1979), and Elmore and LaPointe (1974). No studies were located which found evidence of differences in student evaluations of teachers based on student sex.

With regard to student age, maturity, and/or class level, the majority of the research has failed to find that these variables influence student evaluations of teachers. Much of the research that has been conducted on these variables has addressed whether students possess the maturity and foresight to appreciate the value of the instruction they receive. This research has therefore tended to look at correlations between student ratings at the end of a course with retrospective ratings by the same students after a period of time thought sufficient to allow students time to discover the value of the instruction they have received.

In a cross-sectional study, Centra (1974) found "substantial agreement between current students and alumni (of five years) regarding who have been effective or ineffective teachers" (p. 321). Feldman (1989) conducted a major review and synthesis of research into the reliability of student evaluations and found substantial similarity between ratings by current students and ratings by former students.

In a longitudinal study by Overall and Marsh (1980), the stability of ratings (i.e.,
correlation between ratings and retrospective ratings) was .83. They concluded that "these findings demonstrate that students' evaluations collected at the end of a course are remarkably similar to the retrospective ratings provided by the same students several years later" (Overall & Marsh, 1980, p. 321). Marsh and Overall (1981) also described these findings as demonstrating a correlation substantial enough to demonstrate reliability in the form of stability.

Aleamoni (1987) reported similar results supporting the reliability of student evaluations based both on consistency and on stability. Centra (1993) reviewed research on "student characteristics studied for possible biasing effects on their ratings" (p. 72) and found no research which shows a significant impact of such characteristics on student ratings. Finally, Arden (1989) reported that student evaluations of teaching are unaffected by age, grade point average, and class level/year in school.

**Course Characteristics**

A number of course characteristics have been investigated for potential biasing effect on student evaluations of teaching effectiveness. These include class size, subject matter, class level (graduate or undergraduate), class type (required or elective), and level of difficulty (Centra, 1993).

As discussed in the section on reliability, the research has generally demonstrated an influence of class size. Specifically, "smaller classes get higher ratings in the dimensions of instructional rapport and interaction with students (Marsh, 1987)" (Centra, 1993, p. 66). As a general finding within this research, interrater reliability of student
evaluations appears to vary not only with the instrument being used but also with the class size.

For example, Marsh (1984) summarized the interrater reliability of the SEEQ as being .20 when referring to the degree of agreement between any two students in a class (randomly selected) and as ranging from .60 for a class size of 5 to .95 for a class size of 50 when class-average responses are considered. Similar interrater reliability figures are cited by Cashin (1988) in reference to the IDEA system: "10 raters = .69; 20 raters = .81; 40 raters = .89" (p. 1). These data indicate a generally acceptable degree of consistency given class sizes of at least 15. Centra (1993) also addresses situations in which one wishes to utilize student evaluations in classes with fewer than 15 students: "Because of the differences in ratings by class size, both the SIR and IDEA systems make comparison data available" (p. 67).

With regard to the potential influence of subject matter on student evaluations, researchers have sought to determine whether various academic areas are associated with higher or lower ratings of instruction. Cashin (1990a) utilized data from the national data banks for both the SIR and IDEA and found that "students do rate different academic fields differently" (p. 113). For example, Cashin (1990a) found that the highest rate group "tends to consist of the arts and humanities" whereas the lowest rated group "tends to consist mostly of business, economics, computer science, math, physical sciences, and engineering" and the "biological sciences and health and other professions tend to fall somewhere in between" (p. 117).

Because of this, Centra (1993) suggested that universities do research within their
institution to determine whether similar differences exist. As Neumann and Neumann (1983) pointed out, if such differences do exist, "the methodological implication is quite clear: studies of students' evaluations of instruction ought to be conducted at the program or department level and ought to avoid the aggregate all-university level" (p. 331).

Regarding the potentially biasing effect of class level, research remains scanty: this potential bias was not present in Centra's (1993) review of such research. Although 80 percent of studies on student evaluations have been conducted in higher education (Finley & Crawley, 1993), relatively few have addressed any differences which might exist between the graduate and undergraduate levels. As Koch (1981) observed, "most of the studies in the literature have been restricted to the ratings of instruction given by undergraduate students, rather than graduate or professional area students" (p. 1). Those studies which have investigated student evaluations at both levels have yielded somewhat mixed results.

Marsh (1982) investigated the validity of student evaluations at both the graduate and undergraduate levels and reported that his "findings...support the validity of student ratings for both graduate and undergraduate courses" (p. 264). Similarly, Mannan and Traicoff (1976) found that graduates and undergraduates agreed on the characteristics most important for effective teaching. Koch's (1981) study, conducted in an Educational Psychology Department, found that "approximately the same dimensions of instructional ratings were important to graduate students as have previously been found for undergraduate ratings of courses and their instructors" (p. 9). Finally, although they
found slight differences due to course level, Marsh and Overall (1981) stated that their "results showed variance attributable to the specific instructor was much larger than that due either to course level or course type" (p. 103).

Other researchers have found slight differences due to course level, however. The only study found to demonstrate quantitative differences was conducted by Marsh and Bailey (1993). They reported that "graduate level courses are evaluated somewhat more favorably than undergraduate courses" (Marsh & Bailey, 1993, p. 7). Several studies found qualitative differences between the characteristics deemed important for effective teaching at the graduate and undergraduate levels. Mazer (1977) did a factor analysis of student ratings and found that the Rapport-Responsiveness factor "account[ed] for 39 percent of the commonality" whereas the Evaluation Process and the Instructional Merit factors accounted for only 9 percent and 8 percent respectively. He indicated that "the emergence of Responsiveness-Rapport as the paramount factor in the matrix was unanticipated" and suggested that "the salience of the factor in the present study...may distinguish graduate students from undergraduates with respect to expectations and the major criteria upon which they quality of instruction is assessed" (p. 9). More recently, Ogden, Chapman and Doak (1994) also conducted a similar factor analysis. They found slight differences: "By far, the most important characteristic cited by undergraduate students, is understanding. The most important characteristic, as cited by graduate students, is caring" (Ogden, Chapman & Doak, 1994, p. 8).

Finally, Smith and Cranton (1992) also found differences in characteristics associated with teaching effectiveness at the graduate and undergraduate levels, with
graduate students rating "Interest and Atmosphere" as most important and undergraduates rating "Organization and Clarity" as most important (p. 761). However, in interpreting these results, the researchers advised the reader to not use these results as evidence against the validity of student evaluations. "...It is reasonable to expect that students would perceive different teaching behaviors to be effective in varying settings and that such ratings are not indicative of a lack of validity of the ratings" (Smith & Cranton, 1992, p. 749).

With regard to the type of class, researchers tend to agree that students rate electives and classes in their major more highly than they rate required classes (Arden, 1989; Centra, 1993; Marsh, 1987; Rutland, 1990). Centra (1993) indicated that "although the students' prior subject interest probably affects course ratings more than it does teacher ratings, some rating systems such as IDEA attempt to take it into account in their instructor reports" (p. 71).

The final course characteristic often thought to bias student ratings of teaching effectiveness is the level of difficulty of the class. Teachers frequently assume that more difficult courses are rated more poorly (Centra, 1993; Cruse, 1987). However, according to research conducted by Marsh (1987), the opposite was true! Centra (1993) summarized Marsh's findings: "the teachers who rated more highly gave more work or were believed to teach more difficult courses" (p. 72).

**Instructor Characteristics**

There has been a substantial amount of research conducted on the impact of
instructor characteristics on student evaluations of teaching effectiveness. Although a thorough review of all of this research is beyond the scope of this inquiry, it is important to address the nature and results of the investigations which have been conducted. This research has tended to examine the impact of one or more of the following instructor characteristics on student evaluations of teaching effectiveness: instructor sex, instructor gender-role orientation, instructor personality, and instructor teaching experience. The research conducted on student evaluations of professors can be divided into three types of studies: analogue studies in which students evaluate hypothetical professors as described in vignettes or written scenarios, actual studies in which students have rated professors after having had them for an actual class in college, and studies in which the teaching opinions or behaviors of instructors are compared.

The findings from all three types of studies have yielded conflicting findings regarding the impact of instructor sex and gender-role orientation on student perceptions of a professor's competence. A number of studies have found student evaluations to be affected by the sex or gender-role orientation of the instructor (Abramson, Goldberg, Greenberg, & Abramson, 1977; Basow & Howe, 1987; Basow & Silberg, 1987; Bennett, 1982; Goldberg & Callahan, 1991; Harris, 1976; Hearn, 1985; Kanekar, 1990; Kaschak, 1978, 1981; Kierstead, D'Agostino, & Dill 1988; Martin, 1984; Minner, 1988; Morrison, et al., 1992; Ogden, Chapman, & Doak, 1994; Sidanius and Crane, 1989) whereas only one study was located which found no differences attributable to the sex or gender-role orientation of the instructor (Elmore & LaPointe, 1974). Two findings by Basow and Silberg characterize the findings of the majority of this research. First, Basow and
Silberg (1987) found that "less favorable ratings of women are most likely to occur when women are not seen as fitting gender stereotypes, in this case by participating in a sex-atypical profession" (p. 312). Second, "because of gender stereotypes, female professors may be expected to be more accessible to students than are male professors" (p. 312).

Although the majority of this research has found males and masculine or androgynous gender-role oriented instructors to be rated higher than females or feminine or undifferentiated gender-role oriented instructors, the results continue to be considered inconclusive. Centra (1979), in acknowledging the lack of conclusive evidence in either direction, suggested that any bias which does exist is of negligible significance: "Male and female teachers are occasionally rated differently, but the differences do not have much practical significance" (p. 33). Likewise, Cashin (1990b) stated that instructor sex should not be given "undue weight" in the interpretation of student ratings except when "the instructor provides evidence in his or her self-report for the influence of [this] variable, or if you or others have such evidence" (p. 2).

Nonetheless, females in academe continue to allege the existence of bias which influences not only student evaluations but also results in women being segregated by discipline and by institutional type; to be disproportionately represented at lower ranks; to get promoted at a slower rate than their male colleagues; to participate less in governance and administration; and to be compensated at a rate that averaged only 85 percent of that of their male colleagues. (Goodwin & Stevens, 1993, p. 167)

Another instructor characteristic commonly alleged to influence student ratings of teaching effectiveness is related to personality and charisma. Research on the potential bias of personality and charisma on student evaluations is perhaps best
highlighted by the "Dr. Fox" studies. According to Marsh (1987), "the Dr. Fox effect is defined as the overriding influence of instructor expressiveness on student evaluations" (p. 15). Cruse (1987) elaborated on the origin of references to the "Dr. Fox" effect:

The original Dr. Fox (Naftulin, Ware, and Donnelly, 1973) was presented as a professor; he was in reality an actor who presented little or no subject matter information, lectured with charm and charisma, but used neologisms, spoke double talk, and presented nonsense. Favorable ratings of the actor led Naftulin et al. (1973) to characterize the effect as educational seduction. (pp. 726-727)

Since that study, a number of researchers have found evidence for the biasing influence of personality and charisma. These include Abrami, Leventhal, and Perry (1982), Meier and Feldhusen, (1979), Murray (1983) and Murray and Lawrence (1980). Most recently, Cruse (1987) asserted that instructor personality and charisma have significant influence on student evaluations of teaching and cited the low correlation between student grades (a measure of student learning) and student ratings of the teacher as evidence for bias. Summarizing his stance, Cruse suggested that,

Given the Classical conception of the good professor, one can see that developing charismatic talents primarily to please students is undesirable, and suggesting ways to spuriously improve ratings is loathsome. However, administrations that insist on using student opinions in faculty evaluations may eventually produce the charismatic features the ratings are most sensitive to. (p. 735)

Other researchers, however, have found little evidence for the existence of what has become known as the "Dr. Fox" effect. Aleamoni (1987) reported that, although this represents a "typical faculty concern" (p. 25), students are not easily fooled and can, indeed, discriminate between personality and content related teaching characteristics. Similarly, Marsh (1987) collaborated with one of the original "Dr. Fox" researchers
(Marsh & Ware, 1982; cited in Marsh 1987) to reanalyze data collected in some of the original studies which found support for the Dr. Fox effect. In reanalyzing this data, the researchers used a "factor analysis of the rating instrument" and found that "the instructor expressiveness manipulation only affected rating of Instructor Enthusiasm, the factor most logically related to that manipulation" (Marsh, 1987, p. 15).

A final area of instructor characteristics frequently researched as a potential source of bias is years of experience. Some research in this area has suggested that there is a small impact of experience on student ratings (Centra, 1979). According to Cross (1988), Centra and Creech (1976; cited in Cross, 1988) found that "teachers with three to twelve years of experience are rated somewhat higher than those with either more or less teaching experience..." (p. 8). Kinney and Smith (1992) also referred to this study in their review of literature; the results of their study, however, suggested that the decline in ratings after the 12th year of teaching exists only in the hard sciences whereas ratings for professors in the humanities and social sciences continue to improve up until retirement age. They offered the following explanation for these findings:

There appears to be adverse self-selection in both the humanities and the physical and biological sciences. The most effective teachers retire early so that the remaining pool of professors will be of lower average effectiveness rating. In the social sciences, in contrast, there is favorable self-selection, as the instructors, rated as less effective, choose to retire early. However, the impact is small in all three discipline groups. (p. 299)

Marsh and Hocevar (1991) also provided evidence suggesting that ratings of teaching effectiveness remain stable over time regardless of teaching experience. Using the SEEQ, Marsh and Hocevar (1991) conducted a longitudinal study of 195 teachers over
a 13 year period and found "almost no changes over time for any of the nine content-specific dimensions, the overall course rating, or the overall instructor rating. These findings were consistent for teachers who had little, moderate, or substantial amounts of teaching experience at the start of the study" (p. 303).

To summarize the findings with regard to potential biases which threaten the validity, the research on student, course, and instructor characteristics has been largely mixed. With respect to students, an interaction effect appears to exist between student and instructor sex/gender-role orientation but is not yet inarguably documented in the research. With regard to course characteristics, class size, subject matter, and type of class all appear to have a small but clear impact on student evaluations. With respect to instructor characteristics, there continues to exist a heated debate about the influence of sex and gender-role orientation. Although Marsh (1987) argued persuasively that student evaluations of teaching effectiveness are probably "the most thoroughly studied of all forms of personnel evaluation, and one of the best in terms of being supported by empirical research" (p. 369), the questions about the potential bias of sex and gender-role orientation seem to remain (Bacchi, 1993; Basow & Silberg, 1987; Coufal & Hines, 1976; Cruse, 1987; Gerlin, 1994; Hensel, 1990; Kaschak, 1978, 1981; Kessler-Harris, 1985; Kierstead, D'Agostino, & Dill, 1988; Maitland, 1990; Sandler & Hall, 1986; Sidanius & Crane, 1989). To examine the influence of these two variables on student evaluations of teaching effectiveness in this department is the purpose of this study. The following section will therefore address in more detail previous research on this topic.
Sex, Gender-Role Orientation, and Student Evaluations

The research conducted on student evaluations of professors can be divided into two types of studies: analogue studies in which students evaluate hypothetical professors as described in vignettes or written scenarios and actual studies in which students have rated professors after having had them for an actual class in college. The findings from both types of studies have yielded conflicting findings regarding the impact of sex and gender stereotypes on student perceptions of a professor's competence.

A number of studies have found males to be rated as more competent and effective as instructors than equally qualified females. In a study of how students rated their actual professors, Basow and Silberg (1987) found that both male and female students gave significantly poorer ratings to their female professors, even when the professors were matched in terms of course division, years of teaching experience, and tenure status. Factors which seemed to influence the differential rating of male and female professors included the professor's sex, the student's sex (male students rated female professors lower than did female students), the professor's behavioral traits (instrumental/active vs. expressive/ nurturant), the subject being taught and the student's major. Two findings are especially applicable to this dissertation. First, Basow and Silberg (1987) found that "less favorable ratings of women are most likely to occur when women are not seen as fitting gender stereotypes, in this case by participating in a sex-atypical profession" (p. 312). Second, "because of gender stereotypes, female professors may be expected to be more accessible to students than are male professors" (p. 312).
Sidanius and Crane (1989) also found that male professors were rated higher than were female professors in a study where students rated actual professors they had for a class. Specifically, they found that "male faculty were given higher evaluations on global teacher effectiveness and academic competence than female faculty" (p. 174). In examining the literature, Sidanius and Crane offered two explanations for the seeming disparity of results. They suggested that 2 Factors might be contributing to the mixed results. First, the amount of information about the professor available to the rater seems to influence the results. Specifically, the less information available, the more salient the sex of the professor seems to be. This has served as the rationale for the decision to utilize actual students in this study rather than using an analogue design. Second, perceptions of a professor's competence seem to be affected by the extent to which his or her behavior violates sex stereotypes. As previously discussed, there does seem to be a "narrow band of acceptable behaviors" available to women in the corporate world (Morrison, et al., 1992, p. 47).

This second factor is also consistent with findings by Martin (1984) that students seem to expect female professors to demonstrate both a high level of feminine warmth and a high level of masculine competence in order to be viewed as equal to male professors. Males, on the other hand, do not seem to be expected to be warm or nurturant. Their ratings have been consistently based only on their professional competence.

Kierstead, D'Agostino, and Dill (1988) also found that female instructors must "excel in both stereotypically masculine (e.g., competence) and feminine (e.g., warmth)
domains" (p. 342) in order to be viewed as equally effective. These findings suggest that female professors must meet both the masculine and the feminine standards in order to be accepted. "Taken as a whole, these results suggest that if female instructors want to obtain high student ratings, they must be not only highly competent with regard to factors directly related to teaching but also careful to act in accordance with traditional sex role expectations" (Kierstead, et al., 1988, p. 344).

Finally, Kaschak conducted two analogue studies on sex bias in student evaluations. In her 1978 study, Kaschak found that female students rated female professors as equally competent but less powerful than male professors. "Male students, however, were consistently biased against female professors or in favor of male professors" (Kaschak, 1981, p. 768). In her 1981 study, Kaschak introduced into the vignettes a description of each professor as being "award-winning." When this piece of information was included, both male and female students rated the male professors as being both more powerful and more effective. An analysis of the results demonstrated that "even when males and females are award-winning university professors, they are judged to have achieved this outstanding success as a result of stereotyped masculine and feminine characteristics respectively" (Kaschak, 1981, p. 771). When females achieve, the accomplishment tends to be attributed to stereotypically female behavior such as warmth or to external factors such as luck. For males, a comparable achievement tends to be attributed to skill or competence, both internal qualities that are also stereotypically male.

In each of these studies, therefore, female professors seem to be evaluated not
only in terms of their level of professional competence, but also in terms of their adherence to sex-role stereotypes. These stereotypes seem to be related to other research which has failed to find any difference between evaluations of male and female instructors but has found a difference between gender related characteristics.

One such study, using an analogue design in which students rated a profile of a professor, was conducted by Basow and Howe (1987). The results indicated that the androgynous professor, high in both masculine and feminine characteristics, was rated the highest regardless of the sex of the professor. However, as the authors acknowledged in their discussion, one of the reasons for this finding was that the design was confounded. The profile of the androgynous professor contained the most information (all positive) of all of the profiles. As Sidanius and Crane (1989) reported, factors such as sex and gender-role orientation can be expected to be more salient when the rater has a lesser amount of information. It is therefore difficult to determine the cause of the difference found in the study by Basow and Howe (1987).

Harris (1975) also conducted an analogue study and found no difference between student evaluations of male and female instructors. However, the study did provide evidence that instructors employing a masculine style of teaching were consistently ranked higher than those using a feminine style of teaching. Harris did another study in 1976, pointing to the need to examine the confounding of sex with gender-stereotypic traits. "The most striking finding [was] the overwhelming tendency for a teacher described in masculine stereo-typed terms to be rated more positively on all variables [including competence] except warmth than a femine [sic] teacher" (Harris, 1976, p. 19).
Therefore, this study also suggests that the differences in evaluation may have more to do with a teacher's gender stereotypic style than with a teacher's sex. It also clearly shows that the masculine items were considered more important to a professor's competence than were the feminine items.

One study which found "no significant or practical differences between ratings of male and female faculty or between male and female students' ratings of faculty" was conducted by Elmore and LaPointe (1974, p. 388). This study involved actual student ratings of professors. These researchers found virtually no differences across sex with the exception of two items. Females were rated higher on the item inquiring about the prompt return of tests and assignments and males were rated higher on speaking understandably. No difference was found on the item labeled "in general, taught the class effectively" (Elmore & LaPointe, 1974, p. 388).

Finally, three studies were found which indicated that female professors had received more positive evaluations. Kanekar (1990) researched student evaluations of professors in Bombay and found that the female professors were rated as "more respectable" but that this higher ranking was limited to the lower ranked disciplines. Kanekar suggested that the results may have been confounded by the fact that 90% of the students in Bombay are female and that the brightest and most talented females go into college teaching in Bombay whereas the brightest males go into more profitable professions.

Bennett (1982) also found that female professors received higher ratings from their actual students. However, the adherence to gender stereotypes one again seems to
have been a contributing factor, suggesting that females must excel not only in competence areas but also in the feminine areas of warmth and nurturance. Bennett elaborated on this phenomenon:

This study does suggest that students are less tolerant of female instructors in a number of respects. For example, women not perceived as especially charismatic, experienced, and professional in instructional style are unlikely to be accepted as offering authoritatively balanced instruction. Further, students clearly demand a higher standard of formal preparation and organization from female instructors. Therefore, in the institution studied, women invested considerably in both the formal aspects of teaching and the informal aspects of the instructor role in order to earn parity with their male colleagues. (1982, p. 178)

Bennett's results seem therefore to be a product of an extraordinary amount of effort on the part of the female professors to be recognized as equal or better than their male counterparts.

Finally, a study by Abramson, Goldberg, Greenberg, and Abramsom (1977) found evidence that females are sometimes rated more positively than males when in actuality the performance was equal. They coined a term for this effect: the "talking platypus phenomenon." According to these authors, the term describes an instance in which "an individual achieves a level of success not anticipated [and the] achievement is magnified rather than diminished. After all, it matters little what the platypus says, the wonder is that it can say anything at all" (1977, p. 114).

In this researcher's opinion, the idea of the talking platypus phenomenon being applied to women in the workplace reflects the ultimate bias against women. To assume that a level of success is magnified because it is a wonder that women can achieve anything at all is offensive. However, it also seems to be rare. The literature reviewed
here seems to show that, in general, women are expected to do more to be evaluated as equal to males and that a failure to adhere to gender stereotypes results in the devaluation of female professors' level of competence.
CHAPTER III

DESIGN

Introduction

This study sought to contribute to the literature and research pertaining to the experiences and evaluations of professors in higher education by studying student evaluations of professors in the fields of counselor education and counseling psychology. More specifically, this study examined a population of graduate students enrolled in master or doctoral level classes in the field of counselor education or counseling psychology and was designed to determine the impact of the sex of the students and of the sex and gender-role orientation of the professor as perceived by the students on the students' evaluations of the professor's competence.

Population and Sample

For the purposes of this study, the population was defined as students in the Counselor Education and Counseling Psychology (CECP) Department at a large university in the Midwest. Because this study investigated the student evaluations of professors in the department, the participants included in the sample were the actual students who participated in the evaluation procedure. This sample included all students enrolled in the CECP department in the Spring, Summer, and Fall of 1995 who
completed the evaluation process.

All students enrolled in a "content" course taught by a full-time professor were invited to participate in this process. For the purposes of this study, "practicum" classes were not surveyed. This is not only because of the fact that this particular department utilizes adjunct instructors for the practica but also because the role of a practicum instructor may be one of "supervision" rather than "teaching." Students were allowed to participate in this study only once in order to avoid analytical problems related to repeated measurements.

A total of 276 students chose to participate in this research. From these students, 249 useable data sets were received. The 249 students completing these useable data sets represent the sample in this study. The students were drawn from a total of 25 separate classes, 13 of which were taught by female professors and 12 of which were taught by male professors.

The plan for this study was approved by the Human Subjects Institutional Review Board (HSIRB) at Western Michigan University. The letter of approval may be found in Appendix D.

Instrumentation

Introduction

Students participating in this study completed a set of three instruments in the following specified order: the Instructional Development and Effectiveness Assessment
(IDEA) System (short form), a modification of the Bem Sex-Role Inventory to describe the professor, and a questionnaire designed by the researcher. In order to allow for proper scoring of the IDEA ratings of professor competence, it was also necessary to collect information from professors regarding the relative importance of each of 10 types of student progress. Professors therefore completed an instrument associated with the Instructional Development and Effectiveness (IDEA) System designed for professors to assign weights to each of 10 student progress objectives.

**Instructional Development and Effectiveness Assessment System**

The Instructional Development and Effectiveness Assessment (IDEA) System is a widely utilized and well studied instrument (Cashin, 1990b; Centra, 1993). The 1988 short version of this instrument was used. It is a 14 item self-report inventory which inquires about the students' reactions to the instructor and to the course and which also elicits information regarding the students' perceptions of their progress with regard to a wide range of instructional goals.

Cashin (1988) reported that the IDEA System's interrater reliability varies depending upon class size and provides a generally acceptable degree of consistency given class sizes of at least 15. Specifically, Cashin (1988) cited the following interrater reliability figures: "10 raters = .69; 20 raters = .81; 40 raters = .89" (p.1).

As discussed earlier, the validity of the IDEA instrument (as well as all other instruments involving student evaluation of instructor effectiveness) has been especially difficult to research and document. This is because there are no universally accepted
criteria for what constitutes effective teaching (Cashin, 1988; Feldman, 1988; Marsh, 1984) and therefore no firm criteria against which to compare the ratings by students. As discussed earlier, research intended to measure the validity of student evaluation instruments, such as the IDEA System, has therefore tended to compare student evaluations of professors to a measure of student learning or to other evaluations of teacher effectiveness which are assumed to be valid. These other sources have included instructor self-evaluation, peer/colleague evaluations, administrator evaluations, and alumni evaluations.

The developers of the IDEA System (Hoyt & Cashin, 1977) have perhaps been the most enthusiastic supporters of the use of student learning as the criterion for effective teaching. Their argument has been that the logical result of effective teaching is student learning. Although Marsh (1984) identified student learning as "the most widely accepted criterion of effective teaching," (p. 720) the IDEA System is the only widely used student rating instrument which employs student learning as the major criterion for teaching effectiveness. The IDEA System is therefore unique as a student evaluation instrument because of its use of student learning as the primary criterion for professor effectiveness and was chosen for this feature.

This instrument was utilized to ascertain the students' evaluation of the professor's competence. Consistent with the literature on the use of evaluations for summative purposes, only global items were used in this study. Items 1-10 represented student self-reported progress and were weighted and averaged. This average served as the first dependent variable. Items 12 and 14 both focused on descriptions of the
instructor and were averaged to yield the second dependent variable. A copy of the letter granting permission to utilize this instrument is included in Appendix C.

**Modified Bem Sex-Role Inventory**

The Bem Sex-Role Inventory (BSRI) is "designed to categorize individuals according to their sex role as a function of the degree to which they identify with an array of gender-typed attributes" (Bieger, 1985, p. 51). This instrument is unique in its approach to measuring masculinity and femininity because it does not treat these characteristics as opposite or bipolar traits but instead treats femininity and masculinity as two independent dimensions rather than two ends of a single dimension, thereby enabling a person to indicate whether she or he is high on both dimensions ("androgynous"), low on both dimensions ("undifferentiated"), or high on one dimension but low on the other ("feminine" or "masculine"). (Bem, 1981, p. 4)

A modified version of the short form of the instrument was used. This is a 30 item inventory which asks the student to rate the degree to which each item is descriptive of the professor on a 7 point Likert type scale.

Wheeless and Potorti (1989) reported using a similar modification in order to elicit information regarding student perceptions of the professor's gender-role orientation. The results of their study "provide support for the examination of others' assessments of teacher personality characteristics instead of self-assessment" (Wheeless & Potorti, 1989, p. 262). Feldman (1986) also argued against using teacher self-report of personality traits and advocated the use of student perceptions of personality traits instead. Specifically, Feldman's meta-analysis (1986) compared correlations between
measures of teacher effectiveness and teacher self-report of personality characteristics with correlations between measures of teacher effectiveness and teacher personality characteristics. Only 4 of the 14 correlations using teacher self-report of personality characteristics were significant, whereas 11 of the 14 correlations using student perceptions of teacher personality traits were significant. Other researchers have also suggested the importance of determining the relationship between student perceptions of teacher personality traits and their satisfaction with the instruction (Basow & Distenfeld, 1985). It is for this reason that this study used a modified version of the BSRI to ascertain professor gender-role orientation as perceived by students.

The Modified BSRI yielded scores on the Feminine Scale and the Masculine Scale. These scores were then combined and analyzed with a split-half median method to yield the student perception of the professor's gender-role orientation as masculine, feminine, androgynous or undifferentiated.

Internal consistency (alpha) coefficients for the BSRI range from .75 to .90 (Bieger, 1985). Test-retest reliability coefficients ranged from .76 to .94 (Bieger, 1985). The correlation between the BSRI and the BSRI Short Form ranged from +.85 to +.94. Construct validity has been established based on "the quality and quantity of the experiments conducted using the instrument and on the extent of convergence among the results and findings of the various experiments" (Bieger, 1985, p. 54). Taylor (1984) also found evidence to support the concurrent validity of the BSRI by examining the relationship between masculinity, instrumentality, and instrumental behavior and between femininity, emotional expressiveness, and expressive behavior. Because this instrument
purports to measure gender-role orientation based on a unique theoretical perspective (as described above), it is not possible to obtain validity coefficients by correlating the results of the BSRI with other instruments designed to measure masculinity or femininity. A copy of the letter granting permission to utilize this instrument is included in Appendix E.

**Student Questionnaire**

A questionnaire designed by the researcher was used to obtain information on the student's rating of the quality of the course and to elicit demographic information regarding the student. After asking the student to rate the quality of the course on a scale of 0-4, this instrument also asked the students to specify their sex, the sex of the professor, the program in which they are enrolled, their graduate school status (full or part time), and their employment status. Although some of this data were not necessary to the currently proposed analyses, they may be helpful in future research. This instrument is included in Appendix F.

**IDEA Faculty Information Form**

The questionnaire is an integral part of the IDEA rating system. In order to assign appropriate weights to the learning objectives for students, the system requires professors to provide information regarding the relative importance of each of 10 objectives. These ratings were then used to assign weights to each of the objectives on which the students rated their own progress. Objectives described by professors as being
of "minor importance" received a weight of 0. Objectives described by professors as being "important" received a weight of 1. Objectives described by professors as being "essential" received a weight of 2. This instrument is included in Appendix G.

Preparation of Instrumentation

For ease of administration, the instrument questions were retyped on separate pages and stapled together. This allowed the student to complete the IDEA rating form first, the modified BSRI second, and the student questionnaire third. The instruments for professors (the Faculty Information Form and the BSRI) were also retyped and stapled together to allow them to complete the instruments in a specified order. The students and professors were asked to mark their responses on an Op-Scan sheet which was then coded only at the end of the data collection period. They were assured that there would be no pre-coding of these answer sheets in order to ensure the anonymity of the students and the confidentiality of the professors.

Methodology

Preparation

The researcher first met with the department chair to request permission to speak at a faculty meeting about her research proposal. She then attended this meeting to briefly describe the nature of her research and to seek consent from each full-time professor to conduct the research in their classes. (The researcher sought permission to
collect data only in those classes taught by a tenure-track, full-time faculty member. Courses taught in a practicum format were also excluded because the professor's role in these courses may be considered more a role of supervisor than of teacher.) One hundred percent of the faculty agreed to participate by allowing the researcher to conduct the research in their classes and by agreeing to complete the forms for the professors.

A system was then developed to ensure a balance of courses taught by males and females. Specifically, a system was devised within which all non-practicum courses taught by full-time professors in the department were arbitrarily assigned an identification number. Within this system, in the event that the number of courses taught by men were different than the number of courses taught by women, a table of random numbers was used to determine which classes would be invited to participate. The procedure was followed independently during each semester of data collection.

All instrumentation, along with the necessary approval form, was then submitted to the Human Subjects Institutional Review Board for permission to conduct the research at the university. Finally, the researcher arranged a schedule with each instructor to administer the instruments. It was decided that the time of data collection would be approximately three-fourths of the way through the semester and would be at least one class meeting prior to the date on which the department course evaluations were scheduled to be conducted.

**Data Collection**

Over the course of three semesters, a total of 25 courses were surveyed over 3
semesters, with 13 of the courses taught by female professors and 12 taught by male professors.

In each class, the Oral Recruitment Script was read by the researcher or a trained representative to the group of potential participants requesting that students participate in the study by completing the set of three instruments. (See Appendix G for a copy of this form.) Potential participants also received a copy of an Informed Consent Form to read. (See Appendix H for a copy of the form.) Potential participants were advised that participation would be completely voluntary and that their responses would be anonymous. They were also assured that the evaluation data yielded by this study would not be made available to the university for use in the tenure and promotion process. In addition, in order to avoid analytical difficulties involving repeated measurements, potential participants were asked to exclude themselves if they had participated in this study in a previous class.

The researcher or trained representative then administered in the classroom the Instructional Development and Effectiveness Assessment instrument, the modified Bem Sex-Role Inventory, and the Student Questionnaire to each of the students who volunteered to participate in the study. The students were instructed to complete the instruments in the order specified above and to not look back to check or alter earlier responses. The researcher or trained representative then collected the completed forms as students completed them in class.

Simultaneously, the researcher administered the IDEA Faculty Information Form and the BSRI Short Form to the professor. The professor's answer sheet was then
placed in a manila envelope with the student answer sheets. The envelope was then labeled "Complete" and the number of student participants in that class was recorded.

**Coding of Response Sheets**

Each Op-Scan answer sheet was then coded with a subject number and a class number for the purposes of analysis. The anonymity of students was thereby ensured by avoiding any pre-coding and by keeping no record of student participants. In order to ensure that professors were also guaranteed confidentiality and that it would not be possible for the researcher to determine the individual evaluations of each professor, the researcher adhered to the following procedures when collecting and coding the data.

1. Administered student and professor instruments which contained no identifying information or codes.

2. Placed all response sheets from a given class in a single manila envelope and mark "completed" along with the number completed by students.

3. Repeated the process for each class, storing all envelopes containing completed response sheets in a single box.

4. After all data had been collected, shuffled envelopes.

5. Marked each envelope with an arbitrary class number (001, 002, 003, etc.).

6. Opened each envelope and coded all response sheets with the class number arbitrarily placed on the envelope.

7. Coded each student response sheet with sequential but arbitrary numbers indicating subject numbers.
Each response sheet was therefore coded with a six digit number for purposes of statistical analysis. However, because these numbers were assigned arbitrarily after the period of data collection, the responses can be considered anonymous.

**Scoring**

All instruments were next scored using an Op-Scan machine. The Response sheets were fed by a trained professional through the Op-Scan machine and the data were then recorded electronically on a computer disk. The data were then transformed using a computer program written by a computer programmer hired by the researcher. This computer program resulted in the computation of all calculations necessary to arrive at each variable necessary for the data analysis. This included the computation of the mean weighted averages of items 1-10 on the IDEA System instruments (to arrive at a single measure of student self-reported progress) and the computation of gender-role orientation using the split-half median method. This output contained data in nine columns as listed in Table 2.

**Data Analysis**

Two types of analyses allowed the researcher to address the five research questions. One type of analysis involved the Analysis of Variance (ANOVA). The second type of analysis involved the comparison of Coefficients of Determination. Each type of analysis was conducted three times: once for each of the three dependent variables.
Table 2
Data Output Columns

<table>
<thead>
<tr>
<th>Column</th>
<th>Variable Type</th>
<th>Variable Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Independent Variable # 1</td>
<td>Sex of Student</td>
</tr>
<tr>
<td>2</td>
<td>Independent Variable # 2</td>
<td>Sex of Professor</td>
</tr>
<tr>
<td>3</td>
<td>Independent Variable # 3</td>
<td>Gender-Role Orientation as Perceived by Students</td>
</tr>
<tr>
<td>4</td>
<td>Independent Variable # 4</td>
<td>Gender-Role Orientation as Self-Reported by Professors</td>
</tr>
<tr>
<td>5</td>
<td>Dependent Variable # 1</td>
<td>Professor Competence Rating # 1: Student Self-Reported Progress</td>
</tr>
<tr>
<td>6</td>
<td>Dependent Variable # 2</td>
<td>Professor Competence Rating #: Student Rating of Instructor</td>
</tr>
<tr>
<td>7</td>
<td>Dependent Variable # 3</td>
<td>Professor Competence Rating #: Student Rating of the Quality of the Course</td>
</tr>
<tr>
<td>8</td>
<td>Student Participant Number</td>
<td>Arbitrarily Assigned</td>
</tr>
</tbody>
</table>

The rationale for conducting three separate ANOVAs as opposed to one Multivariate Analysis of Variance (MANOVA) was that the MANOVA is designed to answer a question in which the researcher was uninterested. Specifically, the MANOVA
overall test is intended to test the significance of a composite score representing the optimal linear combination of the various dependent variables. This question was not of interest in this study.

Instead, the researcher sought to determine whether there were main effects or interaction effects between student sex, professor sex, and professor gender-role orientation as perceived by the student on three separate but related dependent variables reflecting the student's evaluation of professor competence (student progress on course objectives, student rating of instructor effectiveness, and student rating of course quality). To ensure an experimentwise error rate of 0.05, the F from the test of the model significance was compared against 0.0056.

Therefore, in order to test for main effects of sex (of both students and professors) and of gender-role orientation (of professors as perceived by students) and to test for interaction effects, a 2x2x4 analysis of variance (ANOVA) was first utilized. One dependent variable in this study was the weighted average of student self-reported progress with respect to the objectives listed in items 1-10 on the IDEA instrument. A second dependent variable was the average of items 12 and 14 on the IDEA instrument, both of which involve the student's rating of the quality of the instructor. The third dependent variable involved the student's assessment of the quality of the course and was assessed by question 45, located in the Student Questionnaire. Three independent variables were considered: (1) student sex, (2) professor sex, and (3) professor gender-role orientation as perceived by students.

For each of the above analyses, however, the test of model significance suggested
that the 3 factor ANOVA model was not significant. Three separate sets of 2 factor ANOVAs were therefore completed, each one including two of the three independent variables. In testing the model significance of each of the 2 Factor ANOVAs, the model was not significant when including the interaction variables. It was significant, however, when not including the interaction variables. The three sets of 2 factor ANOVAs were therefore computed three times for each of the first four research questions: once for each of the three dependent variables.

Specifically, to test the first hypothesis (that there would be no significant difference between ratings of professor competence by male and female students), three sets of 2 Factor ANOVAs were completed: one using the weighted average of student self-reported progress with respect to the objectives listed in items 1-10 as the dependent variable; one using the student's rating of the quality of the instructor as the dependent variable; and one using the student's assessment of the quality of the course as the dependent variable.

To test the second hypothesis (that male professors would receive higher ratings of competence from students than would female professors), three sets of 2-Factor ANOVAs were also completed: one using the weighted average of student self-reported progress with respect to the objectives listed in items 1-10 as the dependent variable; one using the student's rating of the quality of the instructor as the dependent variable; and one using the student's assessment of the quality of the course as the dependent variable.

To test the third hypothesis (that professors who are perceived as masculine or androgynous would receive higher ratings from students than those perceived as feminine
or undifferentiated), three sets of 2 Factor ANOVAs were again completed: one using
the weighted average of student self-reported progress with respect to the objectives
listed in items 1-10 as the dependent variable; one using the student's rating of the quality
of the instructor as the dependent variable; and one using the student's assessment of the
quality of the course as the dependent variable.

With regard to the fourth hypothesis (that there would be interaction effects
between sex and gender-role orientation), the results of the tests of model significance
for both the 3 Factor and 2 Factor ANOVAs were utilized and yielded a determination
that the models were not significant when the interaction variables were included in the
model.

For each of the above four questions, the ANOVA tested the general null
hypothesis that the means of all groups sampled come from populations with equal mean
student ratings of professor competence on each of the three dependent variable
measures of competence and differed only because of sampling error. A $p = .05$
significance level was chosen for purposes of this study as a widely accepted significance
level for studying human populations. Again, as stated earlier, a $p = .0056$ significance
level was utilized for the tests of model significance in order to ensure that the
experimentwise error rate was no greater than .05.

To test the fifth hypothesis (that gender-role orientation would explain a greater
proportion of the variability in the student ratings of professor competence than would
the sex of the professor or of the student), three 1 Factor ANOVAs were computed.
The coefficients of determination were then computed using information from the
ANOVA summary tables: for example, the coefficient of determination for student sex was computed by dividing the sum of squares for student sex by the sum of squares total. This figure represents the proportion of variability on the dependent variable that can be explained with information on student sex.

Limitations

Because this study was limited to the Counselor Education and Counseling Psychology department at one university in the Midwest, one limitation of the study was its limited generalizability. This study was conducted with the knowledge that it would not be appropriate to make generalizations regarding the impact of professor sex, student sex, or a professor's gender-role orientation on student evaluations of the professor's competence for professors in other departments or at other universities. However, this limited generalizability was acceptable because it was balanced with increased specificity. The results were anticipated to be useful in allowing specific application to the particular department from which data were collected and to therefore facilitate detailed interpretations and further experimentation.

A second limitation stemmed from the fact that the male professors in the department currently have more experience and a higher rank than do the female professors. Ideally, professors would be matched with regard to experience, rank, course title, etc. This was not possible in the department in which this research was conducted. Therefore, the results will need to be interpreted with caution. However, some research has shown that the student evaluations of teaching effectiveness remain stable for
professors regardless of experience. Marsh and Hocevar (1991) studied 195 professors over a period of 13 years.

For both undergraduate and graduate level courses, there were almost no changes over time for any of the nine content-specific dimensions, the overall course rating, or the overall instructor rating. The results were consistent for teachers who had little, moderate, or substantial amounts of teaching experience at the start of the study. (p. 303)

This research therefore suggested that the results of this study are valid despite having been unable to match professors with regard to experience or rank.

Ethical Issues

As indicated in the section on preparation, the researcher received approval from the Human Subjects Institutional Review Board to conduct this research. However, the researcher believed that, ethically, it was also necessary to gain the confidence and permission of the faculty to conduct this research. In order to do so, the researcher first contacted the department chair to express interest in conducting this research. She then met with the entire faculty during a regularly scheduled faculty meeting and explained her research proposal with specific emphasis on the procedures to be used in order to ensure confidentiality regarding the evaluations of individual faculty members. (These procedures were described in detail in the “Coding of Response Sheets” section earlier in this chapter.) Faculty members at this meeting were then invited to ask questions and express concerns. Following the meeting, 100 percent of faculty agreed to participate.
CHAPTER IV

FINDINGS

Introduction

This chapter will begin by describing the quantification, sex, and gender-role orientation characteristics of the participating students and professors. Information regarding the numbers and sex of student participants will first be shared. Information regarding necessary reductions of the sample to deal with inequalities of cell sizes will also be addressed. This will be followed by information on the numbers, sex, and gender-role orientation of the professors who participated. Finally, composite information regarding the participating students, professors and classes will be shared.

Assumptions underlying the analyses will then be explained and discussed with respect to the actual data which were analyzed and with respect to the implications for interpretation of the results.

Next, the chapter will turn to a description of the results. As described in the earlier section on data analysis, the three factor analyses of variance models were found insignificant. Separate 2 Factor analyses, for which the models were significant at the p<.05 level were therefore utilized to examine the data for main effects and interaction effects. For simplicity and ease of reading, the results of these analyses relevant to each of the research questions described in Chapter III will be shared in order of research
question. Appendix I contains the actual output from the 2 Factor ANOVA data analyses and Appendix J contains the actual output from the 1 Factor ANOVA data analyses.

Finally, the chapter will provide a brief compilation of the results.

Description of Participating Students

The sample for this study consisted of students enrolled in the Counselor Education and Counseling Psychology department who participated in this study and provided useable data sets. As shown in Table 3, there were a total of 249 students in the sample. There were a substantially higher percentage of female students than male students, with 73% of the sample consisting of female students. This is in comparison to a population comprised of 76% female students. This distribution of male and female students in the Counselor Education and Counseling Psychology department has remained constant since 1992 (B. Hubbard, personal communication, December 17, 1996).

Description of Participating Professors

As stated earlier, all full-time professors in the Counselor Education and Counseling Psychology Department agreed to participate by allowing the researcher to conduct research in their classes and by completing the IDEA Faculty Information Form and the BSRI Short Form. As shown in Table 4, a total of 12 professors participated, 6 of whom were male and 6 of whom were female.
Table 3
Comparison of Total Student Participants and Population

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
<th></th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Students</td>
<td>182</td>
<td>73</td>
<td></td>
<td>432</td>
<td>77</td>
</tr>
<tr>
<td>Male Students</td>
<td>67</td>
<td>27</td>
<td></td>
<td>136</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>249</td>
<td>100</td>
<td></td>
<td>568</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4
Total Participating Professors by Sex

<table>
<thead>
<tr>
<th>Sex of Professor</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>100</td>
</tr>
</tbody>
</table>
A combining of data contained in Table 3 and Table 4 is provided here in Table 5. As
Table 5 illustrates, male and female students were fairly evenly divided between classes
taught by male professors and classes taught by female professors.

Table 5

Sex of Student and Professor Participants

<table>
<thead>
<tr>
<th></th>
<th>Female Professors</th>
<th>Male Professors</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Female Students</td>
<td>92</td>
<td>37</td>
<td>90</td>
</tr>
<tr>
<td>Male Students</td>
<td>35</td>
<td>14</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>127</td>
<td>51</td>
<td>122</td>
</tr>
</tbody>
</table>

Table 6 describes the gender-role orientation of the participating professors as
perceived by students and as self-reported by the professors. Within this table, the n "as
described by students" indicates the number of times a professor was rated by the 249
students as masculine, feminine, androgynous, or undifferentiated. The GRO "as self-
reported by professors" was recorded for each student response in the sample, yielding
a total of 249 "self-reports" from the 12 professors. This allows for an ease of
comparisons of the student descriptions and the professor self-report. Professor gender-
Table 6

Gender-Role Orientation of Participating Professors: A Comparison of Student Descriptions and Professor Self-Report

<table>
<thead>
<tr>
<th>Gender Role Orientation</th>
<th>As described by Students</th>
<th>As Self-Reported by Professors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Masculine</td>
<td>63</td>
<td>25</td>
</tr>
<tr>
<td>Feminine</td>
<td>60</td>
<td>24</td>
</tr>
<tr>
<td>Androgynous</td>
<td>65</td>
<td>26</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>61</td>
<td>25</td>
</tr>
</tbody>
</table>

Role orientation as described by students represents the third dependent variable utilized in this study and professor gender-role orientation as self-reported by professors represents the fourth dependent variable used in this study.

As shown in this table, students perceived the participating professors as fairly equally divided among masculine, feminine, androgynous, and undifferentiated gender-role orientations. Professor self-report differed in that substantially fewer identified themselves as having a feminine gender-role orientation and substantially more identified themselves as having an androgynous gender-role orientation.
themselves as having an androgynous gender-role orientation.

Data on Reduction of Sample for Statistical Analyses

Although the entire sample consisted of 249 students, the original cell counts were unequal to such an extent that an ANOVA could not be accurately computed, especially given the combination of a sizable inequality of cell sizes and a lack of homogeneity of variance. Some cells were therefore reduced using procedures to randomly remove observations from cells which were overpopulated. These procedures ensured that every participating student had an equal and independent opportunity to remain in the sample of data which was then used for analysis.

Table 7 illustrates the original cell counts and the cell counts after sample reduction for the 2 Factor ANOVA which included professor sex and professor gender-role orientation as perceived by students in the model. This reduction involved the random removal of 10 subjects from the cell at the intersection of male professors and androgynous professors. This resulted in a total of 239 observations for analysis in this ANOVA procedure.

Table 8 illustrates the original cell counts and the cell counts after sample reduction for the 2 Factor ANOVA which included professor sex and student sex in the model. This reduction involved the random removal of 123 subjects from the cells for female students. Although the both the sample and the population being studied consisted of approximately 74% female students, it was necessary to reduce the discrepancy in order to avoid having this inequality result in unequal weighting for this
variable. This reduction resulted in a total of 126 observations for analysis in this ANOVA procedure.

Table 9 illustrates the original cell counts and the cell counts after sample reduction for the 2 Factor ANOVA which included student sex and professor gender-role orientation as perceived by students in the model. This reduction involved the random removal of 123 subjects from the cells for female students. Again, although the both the sample and the population being studied consisted of approximately 73% female students, it was necessary to reduce the discrepancy in order to avoid having this inequality result in unequal weighting for this variable. This reduction resulted in a total of 126 observations for analysis in this ANOVA procedure.

<table>
<thead>
<tr>
<th></th>
<th>Professor Sex</th>
<th>GRO = M</th>
<th>GRO = F</th>
<th>GRO = A</th>
<th>GRO = U</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Cell Counts</td>
<td>Female</td>
<td>36</td>
<td>34</td>
<td>19</td>
<td>38</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>27</td>
<td>26</td>
<td>46</td>
<td>23</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>63</td>
<td>60</td>
<td>65</td>
<td>61</td>
<td>249</td>
</tr>
<tr>
<td>Cell Counts</td>
<td>Female</td>
<td>36</td>
<td>34</td>
<td>19</td>
<td>38</td>
<td>127</td>
</tr>
<tr>
<td>After Reduction</td>
<td>Male</td>
<td>27</td>
<td>26</td>
<td>36</td>
<td>23</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>63</td>
<td>60</td>
<td>55</td>
<td>61</td>
<td>239</td>
</tr>
</tbody>
</table>
Table 8

Original Cell Counts and Cell Counts After Reduction: 2 Factor ANOVAs
With Professor Sex and Student Sex in Model

<table>
<thead>
<tr>
<th></th>
<th>Female Students</th>
<th>Male Students</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Cell Counts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female Profs</td>
<td>92</td>
<td>35</td>
<td>127</td>
</tr>
<tr>
<td>Male Profs</td>
<td>90</td>
<td>32</td>
<td>122</td>
</tr>
<tr>
<td>Totals</td>
<td>182</td>
<td>67</td>
<td>249</td>
</tr>
<tr>
<td>Cell Counts After</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduction</td>
<td>Female Profs</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Male Profs</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>67</td>
<td>126</td>
</tr>
</tbody>
</table>

Description of Classes Surveyed

Data were collected from a total of 25 classes in the Counselor Education and Counseling Psychology Department. Each of these classes was a graduate-level class and each was taught in a lecture format (as opposed to a practicum format). Numbers of participating students in each class ranged from 2 to 24, with the mean number of participants equaling 10. Of the 25 classes in which data were collected, 13 classes were taught by female professors and 12 were taught by male professors. Table 10 summarizes this data.
Table 9

Original Cell Counts and Cell Counts After Reduction: 2 Factor ANOVAs With Student Sex and Professor GRO in Model

<table>
<thead>
<tr>
<th>Student Sex</th>
<th>GRO = M</th>
<th>GRO = F</th>
<th>GRO = A</th>
<th>GRO = U</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Cell Counts</td>
<td>Female</td>
<td>48</td>
<td>42</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>15</td>
<td>18</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>63</td>
<td>60</td>
<td>65</td>
<td>61</td>
</tr>
<tr>
<td>Cell Counts After Reduction</td>
<td>Female</td>
<td>16</td>
<td>12</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>15</td>
<td>18</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>31</td>
<td>30</td>
<td>32</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 10

Total Participating Classes and Total Participating Students in Classes Taught by Male and Female Professors

<table>
<thead>
<tr>
<th>Sex of Professor</th>
<th>Number of Classes</th>
<th>Number of Student Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>13</td>
<td>127</td>
</tr>
<tr>
<td>Male</td>
<td>12</td>
<td>122</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>249</td>
</tr>
</tbody>
</table>
Having provided descriptive information about the participating students, professors, and classes, this chapter now turns to a discussion of the assumptions underlying the data analysis procedures.

Tests of the Assumptions

Three assumptions underlie the analysis of variance (ANOVA) procedures utilized in the data analysis. The first assumption is that there is independence of observations, meaning that "the observations are all independent of one another....and that knowing how one of these observations stands relative to the treatment (or populations) means tells us nothing about the other observation" (Howell, 1992, p. 289). Meeting this assumption was ensured by the original design of the study.

The second assumption is that there is homogeneity of variance, meaning that the variance of each observation is relatively homogenous or similar to the variance of other observations in the sample. This assumption was not met by the sample data and significantly different variances were observable in the data. Therefore, to ensure that the results of the ANOVAs are accurate, two procedures were necessary in order to ensure that the results would be valid.

First, because inequality of cell sizes is especially problematic in situations in which there is not homogeneity of variance, cell sizes were reduced as described earlier. Second, a post hoc analysis using the Welch F statistic (a procedure which doesn't rely on the assumption of homogeneity of variance) was conducted. Use of this procedure to deal with heterogeneity of variances was recommended by Wilcox (1987) and
Tomarken and Serlin (1986). This procedure did indeed result in the same findings with regard to significance. It can therefore be assumed that, despite not meeting the assumption of homogeneity of variance, the results to be described in the upcoming sections are accurate and are protected against an increased probability of a Type I error (Howell, 1992).

Finally, the third assumption is that the data will be normally distributed. An interocular test of the data suggested that this assumption was also not met. In each case, the data manifested a negative skew, meaning that the ratings of professor competence were more heavily concentrated on the high end of the continuum of 0 to 4 than on the low end. However, because the ANOVA procedure is an extremely robust analysis, "the assumptions [of homogeneity of variance and normality of distribution] can be violated with relatively minor effects. This is especially true for the normality assumption" (Howell, 1992).

Thus, despite some difficulties in meeting the assumptions of the ANOVA procedures for data analysis, the use of the Welch F procedure and the reduction of cell sizes are sufficient to render the data analyses useful and the results of the analyses valid. It is to a description of these results that this chapter now turns.

Results

Introduction

In approaching a description of the results, this section will begin by providing
a restatement of the research questions and the hypotheses, both in research form and in null form. This section will then provide the results relevant to each individual research question. These results will then be interpreted and discussed in Chapter V.

Restatement of Research Questions and Hypotheses

Because each research question required three or more analyses of data in order to adequately answer each question and because the analyses of variance each applied to more than one research question, a review of the original research questions, the research hypotheses, and the null hypotheses may increase the ease with which the upcoming results may be understood. Table 11 is therefore provided to restate the research questions along with the associated hypotheses. The hypotheses are presented both in research form and in null form. The research hypotheses consist of the researcher's anticipation of results prior to conducting the actual research. The null hypotheses, on the other hand, are the hypotheses which are being directly tested in each data analysis. The rejection of the null hypothesis may or may not result in support of the research hypotheses; the implication of the results will therefore be discussed in Chapter V.

Results Relevant to Research Question 1

The first research question addressed by this study was "Is there a significant difference between ratings of professor competence by male and female students?" Determining whether a main effect of student sex existed in the sample was necessary for
## Table 11

Restatement of Research Questions and Hypotheses

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Research Hypothesis</th>
<th>Null Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is there a significant difference between ratings of professor competence provided by male and female students?</td>
<td>There will be no significant difference between ratings of professor competence by male and female students.</td>
<td>There will be no significant difference between ratings of professor competence by male and female students.</td>
</tr>
<tr>
<td>Is there a significant difference between ratings received by male professors and female professors?</td>
<td>Male professors will receive higher ratings of competence from students than will female professors.</td>
<td>There will be no significant difference between ratings received by male professors and female professors.</td>
</tr>
<tr>
<td>Is there a significant difference among the ratings received by professors who are perceived as masculine, feminine, androgynous, or undifferentiated by students?</td>
<td>Professors who are perceived as masculine or androgynous will receive higher ratings from students than those who are perceived as feminine or undifferentiated.</td>
<td>There will be no significant difference among the ratings received by professors who are perceived as masculine, feminine, androgynous, or undifferentiated by students.</td>
</tr>
<tr>
<td>Are there interaction effects between student sex and professor sex, between professor sex and gender-role orientation, or between student sex and professor gender-role orientation?</td>
<td>Significant interaction effects will exist. Female professors will receive higher ratings from female students and male professors will receive higher ratings from male students.</td>
<td>There will be no significant interaction effects between student sex and professor sex, between professor sex and gender-role orientation, or between student sex and professor gender-role orientation.</td>
</tr>
</tbody>
</table>
Table 11—Continued

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Research Hypothesis</th>
<th>Null Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does sex or gender-role orientation explain a greater proportion of the variability in the student ratings of professor competence?</td>
<td>Gender-role orientation will explain a greater proportion of the variability in the student ratings of professor competence than would the sex of the professor or student.</td>
<td>There will be no significant difference in the a greater proportion of the variability in the student ratings of professor competence explained by sex and the proportion explained by gender-role orientation.</td>
</tr>
</tbody>
</table>

providing an answer to this question. As shown in Table 11, no main effect was anticipated by the researcher.

Two separate sets of 2 Factor ANOVAs were necessary to examine the data for main effects of student sex: one including both student sex and professor sex and the other including student sex and professor gender-role orientation as perceived by students. Because neither the two nor the three factor ANOVA models were found to be significant when including interactions (e.g., A x B or student sex by professor sex) in the model, it can be concluded that there were no significant interaction effects. The interactions were therefore not included in the 2 Factor ANOVAs described here.

As shown in Table 12, the model for the 2 Factor ANOVAs were not significant when including both student sex and professor sex, regardless of whether the dependent variable was student self-reported progress, student rating of the course, or student
Table 12

Test of Model Significance for 2 Factor ANOVA Including Student Sex and Professor Sex in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Model</td>
<td>2</td>
<td>0.8233</td>
<td>0.65</td>
<td>0.5250</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>123</td>
<td>78.1649</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>78.9882</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Model</td>
<td>2</td>
<td>3.5143</td>
<td>1.26</td>
<td>0.2876</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>123</td>
<td>171.6920</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>175.2063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Model</td>
<td>2</td>
<td>6.6955</td>
<td>2.05</td>
<td>0.1326</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>123</td>
<td>200.5109</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>207.2064</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

rating of the instructor. This indicates that there were no main effects of student sex when both student sex and professor sex were included in the model.

As shown in Table 13, the models for the 2 Factor ANOVAs were all significant when including student sex and professor gender-role orientation as perceived by students in the model. This indicates that further analysis for main effects was warranted. Table 14 shows the ANOVA summary tables for each of the three ANOVAs using this model.

In all cases, female students gave higher ratings of professor competence than did
### Table 13

Test of Model Significance for 2 Factor ANOVA Including Student Sex and Professor Gender-Role Orientation in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Model</td>
<td>4</td>
<td>18.9523</td>
<td>9.55</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>121</td>
<td>60.0359</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>78.9882</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Model</td>
<td>4</td>
<td>41.7374</td>
<td>9.46</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>121</td>
<td>133.4689</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>175.2063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Model</td>
<td>4</td>
<td>41.6034</td>
<td>7.60</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>121</td>
<td>165.6030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>207.2064</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 14

ANOVA Summary Table for 2 Factor ANOVA Including Student Sex and Professor Gender-Role Orientation in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>2.2975</td>
<td>4.63</td>
<td>0.0334</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>18.2014</td>
<td>12.23</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Course Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>2.7687</td>
<td>2.51</td>
<td>0.1157</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>41.3052</td>
<td>12.48</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>1.2054</td>
<td>0.88</td>
<td>0.3499</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>41.6023</td>
<td>10.13</td>
<td>0.0001</td>
<td>YES</td>
</tr>
</tbody>
</table>
male students. However, only in one case was the difference in ratings statistically significant. Stated another way, only one main effect was found. When including both student sex and professor gender-role orientation as perceived by students in the model, there was a main effect of student sex when using student self-reported progress as the dependent variable. Table 15 shows the mean ratings of professor competence provided by male and female students.

To summarize the results relevant to research question 1, female students consistently rated professors more highly than did male students. However, only one statistically significant difference in ratings was found. This main effect was found in the 2 Factor ANOVA including student sex and professor gender-role orientation in the

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Sex of Student</th>
<th>Mean Rating</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Female Male</td>
<td>2.8030</td>
<td>0.0334</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Female Male</td>
<td>3.0875</td>
<td>0.1157</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.7862</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Female Male</td>
<td>3.0353</td>
<td>0.3499</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.8365</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
model and indicated that female students reported having made significantly more progress than did male students. No main effects were indicated in the one factor ANOVAs.

Because only one of the six analyses resulted in a main effect, however, this research study failed to find a consistently significant difference between ratings of professor competence by male and female students. The null hypothesis was therefore not rejected and the results were consistent with the research hypothesis.

Results Relevant to Research Question 2

The second research question addressed by this study was "Is there a significant difference between ratings received by male professors and female professors?" Determining whether a main effect of professor sex existed in the sample was necessary for providing an answer to this question. As shown in Table 11, a main effect in which male professors received higher ratings than female professors was anticipated.

Two separate sets of ANOVAs were necessary to examine the data for main effects of professor sex: one including student sex and the other including professor gender-role orientation as perceived by students. Because neither the two nor the three factor ANOVAs models were found to be significant when including interactions in the model, it can be concluded that there were no significant interaction effects. The interactions were therefore not included in the 2 Factor ANOVAs described here.

As shown in Table 16, the model for the 2 Factor ANOVAs were not significant when including both professor sex and student sex, regardless of whether the dependent
As shown in Table 17, the models for the 2 Factor ANOVAs were all significant when including professor sex and professor gender-role orientation as perceived by students in the model. This indicates that further analysis for main effects was warranted. Table 18 shows the ANOVA summary tables for each of the three ANOVAs using this model.
Table 17

Test of Model Significance for 2 Factor ANOVA Including Professor Sex and Professor Gender-Role Orientation in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Model</td>
<td>4</td>
<td>26.3186</td>
<td>12.48</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>234</td>
<td>123.3842</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>238</td>
<td>149.7028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Model</td>
<td>4</td>
<td>49.1492</td>
<td>10.87</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>234</td>
<td>264.5830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>238</td>
<td>313.7322</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Model</td>
<td>4</td>
<td>71.4187</td>
<td>14.69</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>234</td>
<td>284.3993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>238</td>
<td>355.8180</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 18

ANOVA Summary Table for 2 Factor ANOVA Including Professor Sex and Professor Gender-Role Orientation in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Professor Sex</td>
<td>1</td>
<td>0.3683</td>
<td>0.70</td>
<td>0.4041</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>26.2106</td>
<td>16.57</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Course Rating</td>
<td>Professor Sex</td>
<td>1</td>
<td>2.9576</td>
<td>2.62</td>
<td>0.1072</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>40.4534</td>
<td>11.93</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Professor Sex</td>
<td>1</td>
<td>6.5727</td>
<td>5.41</td>
<td>0.0209</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>56.1250</td>
<td>15.39</td>
<td>0.0001</td>
<td>YES</td>
</tr>
</tbody>
</table>
In all but one case, male professors received higher ratings than did female professors. However, in only one case was the difference in ratings statistically significant. Stated another way, only one main effect was found. When including both professor sex and professor gender-role orientation as perceived by students in the ANOVA model, there was a main effect of professor sex when using the student rating of instructor as the dependent variable. Table 19 shows the mean ratings of professor competence received by male and female professors.

To summarize the results relevant to research question 2, male professors generally received higher ratings from students than did female professors. However, only one statistically significant difference in ratings was found. This main effect was

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Professor Sex</th>
<th>Mean Rating</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Female</td>
<td>2.7578</td>
<td>0.4041</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>2.6774</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Female</td>
<td>2.8760</td>
<td>0.1072</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>3.1039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Female</td>
<td>2.7831</td>
<td>0.0209</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>3.1229</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
found in the 2 Factor ANOVA including professor sex and professor gender-role orientation in the model and indicated that, in this case, students provided statistically higher ratings of the instructor for male professors than for female professors.

Interestingly, a one factor ANOVA yielded two main effects. Male professors received significantly higher ratings than female professors when the dependent variables were ratings of the course and ratings of the instructor respectively. In neither the 2 Factor nor the one factor ANOVAs was there a statistically significant difference in student self-reported progress.

Because only one of the six 2 Factor analyses resulted in a main effect, this research study failed to find a consistently significant difference in the ratings received by male and female professors when including professor gender-role orientation in the model. However, when removing professor gender-role orientation as a source of variability in the ratings of professor competence, two of the three one factor ANOVAs suggested a significant difference in the ratings received by male and female professors. It is therefore only with ambiguity that a decision of whether or not to reject the null hypothesis can be made. The interpretation of these results with respect to the research question will be addressed in greater depth in Chapter V.

**Results Relevant to Research Question 3**

The third research question addressed by this study was "Is there a significant difference among the ratings received by professors who are perceived as masculine, feminine, androgynous, or undifferentiated by students?" Determining whether a main
effect of professor gender-role orientation existed in the sample was necessary for providing an answer to this question. As shown in Table 11, the researcher anticipated a main effect and specifically expected that professors who were perceived as masculine or androgynous would receive ratings that were higher than would professors who were perceived as feminine or undifferentiated.

Two separate sets of 2 Factor ANOVAs were necessary to examine the data for main effects of professor gender-role orientation as perceived by students: one including student sex and the other including professor sex. Because neither the two nor the three factor ANOVAs models were found to be significant when including interactions in the model, it can be concluded that there were no significant interaction effects. The interactions were therefore not included in the 2 Factor ANOVAs described here.

As shown in Table 20, the models for the 2 Factor ANOVAs were all significant when including student sex in the model, regardless of whether the dependent variable was student self-reported progress, student rating of the course, or student rating of the instructor. This indicates that further analysis for main effects was warranted.

Table 21 shows the ANOVA summary tables for each of the three ANOVAs using this model. In all cases, a main effect of professor gender-role orientation as perceived by students was found.

Also, in all cases, the order of rankings (from highest to lowest) received by professors was: androgynous, feminine, masculine, and undifferentiated. Table 22 shows the mean ratings received according to professor gender-role orientation when student sex is included in the model.
### Table 20

Test of Model Significance for 2 Factor ANOVA Including Professor Gender-Role Orientation and Student Sex in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Model</td>
<td>4</td>
<td>18.9523</td>
<td>9.55</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>121</td>
<td>60.0359</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>78.9882</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating</td>
<td>Model</td>
<td>4</td>
<td>41.7374</td>
<td>9.46</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>121</td>
<td>133.4689</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>175.2063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Model</td>
<td>4</td>
<td>41.6034</td>
<td>7.60</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>121</td>
<td>165.6030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>125</td>
<td>207.2064</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 21

ANOVA Summary Table for 2 Factor ANOVA Including Professor Gender-Role Orientation and Student Sex in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>2.2975</td>
<td>4.63</td>
<td>0.0334</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>18.2014</td>
<td>12.23</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Course Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>2.7687</td>
<td>2.51</td>
<td>0.1157</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>41.3052</td>
<td>12.48</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>1.2054</td>
<td>0.88</td>
<td>0.3499</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>Prof. GRO</td>
<td>3</td>
<td>41.6023</td>
<td>10.13</td>
<td>0.0001</td>
<td>YES</td>
</tr>
</tbody>
</table>
Table 22

Mean Ratings of Professor Competence for Professors According to GRO in 2 Factor ANOVAs Including Professor GRO and Student Sex in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Professor GRO</th>
<th>Mean Rating</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Androgynous</td>
<td>3.2097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progress Rating</td>
<td>Feminine</td>
<td>2.7424</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Masculine</td>
<td>2.5574</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undifferentiated</td>
<td>2.1539</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Androgynous</td>
<td>3.7564</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feminine</td>
<td>3.0968</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Masculine</td>
<td>2.7048</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undifferentiated</td>
<td>2.1893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Androgynous</td>
<td>3.6655</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feminine</td>
<td>3.2865</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Masculine</td>
<td>2.6097</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undifferentiated</td>
<td>2.1819</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Secondary analysis of these results was also conducted to determine the significance of differences between each pairing of two gender role orientations. In Table 23, the probability that the ratings received by professors of the two different gender-role orientations was equal is indicated numerically. Any value less than .05 indicates a statistically significant difference.
Table 23

Probability of Statistically Significant Differences Between Ratings of Professor Competence According to Professor Gender-Role Orientation

<table>
<thead>
<tr>
<th>Model Including Professor GRO and Student Sex</th>
<th>Progress Ratings</th>
<th>Course Ratings</th>
<th>Instructor Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>A = F</td>
<td>0.0102*</td>
<td>0.0149*</td>
<td>0.2050</td>
</tr>
<tr>
<td>A = M</td>
<td>0.0004*</td>
<td>0.0001*</td>
<td>0.0005*</td>
</tr>
<tr>
<td>A = U</td>
<td>0.0001*</td>
<td>0.0001*</td>
<td>0.0001*</td>
</tr>
<tr>
<td>F = M</td>
<td>0.3089</td>
<td>0.1490</td>
<td>0.0262*</td>
</tr>
<tr>
<td>F = U</td>
<td>0.0013*</td>
<td>0.0009*</td>
<td>0.0003*</td>
</tr>
<tr>
<td>M = U</td>
<td>0.0239*</td>
<td>0.0522</td>
<td>0.1467</td>
</tr>
</tbody>
</table>

* indicates significance at the p<.05 level

This same analytical procedure was conducted for the 2 Factor ANOVA including both professor gender-role orientation as perceived by students and professor sex in the model. As shown in Table 24, the models for the 2 Factor ANOVAs were all significant when including professor sex in the model, regardless of whether the dependent variable was student self-reported progress, student rating of the course, or student rating of the instructor. This indicates that further analysis for main effects was warranted.
Table 24

Test of Model Significance for 2 Factor ANOVA Including Professor Gender-Role Orientation and Professor Sex in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Model</td>
<td>4</td>
<td>26.3186</td>
<td>12.48</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>234</td>
<td>123.3842</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>238</td>
<td>149.7028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Model</td>
<td>4</td>
<td>49.1492</td>
<td>10.87</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>234</td>
<td>264.5830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>238</td>
<td>313.7322</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Model</td>
<td>4</td>
<td>71.4187</td>
<td>14.69</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>234</td>
<td>284.3993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>238</td>
<td>355.8180</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 25 shows the ANOVA summary tables for each of the three ANOVAs using this model. In all cases, a main effect of professor gender-role orientation as perceived by students was found.

Also, in all cases, the order of rankings (from highest to lowest) received by professors was: androgynous, feminine, masculine, and undifferentiated. Table 26 shows the mean ratings received according to professor gender-role orientation when professor sex is included in the model.

Secondary analysis of these results was also conducted to determine the significance of differences between each pairing of two gender role orientations. In
Table 25
ANOVA Summary Table for 2 Factor ANOVA Including Professor Sex and Professor Gender-Role Orientation in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Professor Sex</td>
<td>1</td>
<td>0.3683</td>
<td>0.70</td>
<td>0.4041</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>GRO</td>
<td>3</td>
<td>26.2106</td>
<td>16.57</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Course Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>2.9576</td>
<td>2.62</td>
<td>0.1072</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>GRO</td>
<td>3</td>
<td>40.4536</td>
<td>11.93</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Student Sex</td>
<td>1</td>
<td>6.5727</td>
<td>5.41</td>
<td>0.0209</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>GRO</td>
<td>3</td>
<td>56.1250</td>
<td>15.39</td>
<td>0.0001</td>
<td>YES</td>
</tr>
</tbody>
</table>

Table 27, the probability that the ratings received by professors of the two different gender-role orientations was equal is indicated numerically. Any value less than .05 indicates a statistically significant difference.

To summarize the results relevant to research question 3, androgynous professors consistently received the highest ratings and were followed, in order of rating, by feminine professors, masculine professors, and finally by undifferentiated professors. In all six sets of the 2 Factor ANOVAs, a main effect of professor gender-role orientation was found. The null hypothesis, which stated that there would be no significant difference but postulated no specific order, was therefore rejected. However, the
Table 26

Mean Ratings of Professor Competence for Professors According to GRO in 2 Factor ANOVAs Including Professor Sex and Professor GRO in Model

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Professor GRO</th>
<th>Mean Rating</th>
<th>p value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progress Rating</td>
<td>Androgynous</td>
<td>3.1678</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Feminine</td>
<td>2.7889</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masculine</td>
<td>2.7029</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undifferentiated</td>
<td>2.2106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Rating</td>
<td>Androgynous</td>
<td>3.5648</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Feminine</td>
<td>3.1485</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masculine</td>
<td>2.8417</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undifferentiated</td>
<td>2.4050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor Rating</td>
<td>Androgynous</td>
<td>3.5748</td>
<td>0.0001</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>Feminine</td>
<td>3.2810</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masculine</td>
<td>2.6195</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undifferentiated</td>
<td>2.3369</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

research hypothesis, which postulated a specific order of rankings, was not supported because the actual order of rankings differed from the order anticipated by the researcher.
Probability of Statistically Significant Differences Between Ratings of Professor Competence According to Professor Gender-Role Orientation

<table>
<thead>
<tr>
<th>Model Including Professor GRO and Professor Sex</th>
<th>Progress Ratings</th>
<th>Course Ratings</th>
<th>Instructor Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>A = F</td>
<td>0.0062*</td>
<td>0.0394*</td>
<td>0.1598</td>
</tr>
<tr>
<td>A = M</td>
<td>0.0007*</td>
<td>0.0003*</td>
<td>0.0001*</td>
</tr>
<tr>
<td>A = U</td>
<td>0.0001*</td>
<td>0.0001*</td>
<td>0.0001*</td>
</tr>
<tr>
<td>F = M</td>
<td>0.5122</td>
<td>0.1110</td>
<td>0.0010*</td>
</tr>
<tr>
<td>F = U</td>
<td>0.0001*</td>
<td>0.0002*</td>
<td>0.0001*</td>
</tr>
<tr>
<td>M = U</td>
<td>0.0002*</td>
<td>0.0233*</td>
<td>0.1511</td>
</tr>
</tbody>
</table>

* indicates significance at the p<.05 level

Results Relevant to Research Question 4

The fourth research question addressed by this study was "Are there interaction effects between student sex and professor sex, between professor sex and gender-role orientation, or between student sex and professor gender-role orientation?" Determining where interaction effects were statistically significant was necessary for providing an answer to this question. As shown in Table 11, the researcher anticipated interaction
answer to this question. As shown in Table 11, the researcher anticipated interaction effects between student sex and professor sex.

This research question was answered by including interaction variables in the three factor and 2 Factor ANOVA models. In each case, the models were found to be nonsignificant when including the interaction variables in the model. It can therefore be concluded that there were no significant interaction effects among student sex, professor sex, and professor gender-role orientation in the study (personal communication, A. Meier, January 31, 1997). This means that the marginal effects of student sex, professor sex, and professor gender-role orientation were consistent across all levels of the other factors.

The null hypothesis was therefore not rejected and the research hypothesis was refuted.

Results Relevant to Research Question 5

The fifth research question addressed by this study was "Does sex or gender-role orientation explain a greater proportion of variability in the student ratings of professor competence?" A comparison of the coefficients of determination was necessary to answer this question. By definition, the coefficient of determination equals the proportion of variability in the dependent variable explained by the independent variable. Calculation of the coefficient of determination is achieved by dividing the sum of squares for a variable by the total sum of squares.

To respond to this question, the one factor ANOVA output was utilized to
calculate the proportion of variability explained by each factor independent of the others.

Table 28 shows the output utilized, the coefficients of determination calculated, and the mean coefficient of determination yielded by each factor: by gender-role orientation as perceived by students, by sex of student, and by sex of professor.

Table 28

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>SS for IV</th>
<th>SS Total</th>
<th>R-Sq</th>
<th>Mean R-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor GRO</td>
<td>Progress</td>
<td>27.8835</td>
<td>152.7634</td>
<td>0.1825</td>
<td>0.1811</td>
</tr>
<tr>
<td></td>
<td>Course</td>
<td>52.5210</td>
<td>322.0000</td>
<td>0.1631</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructor</td>
<td>71.9828</td>
<td>364.3072</td>
<td>0.1976</td>
<td></td>
</tr>
<tr>
<td>Student Sex</td>
<td>Progress</td>
<td>0.7509</td>
<td>78.9883</td>
<td>0.0095</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Course</td>
<td>0.4323</td>
<td>175.2063</td>
<td>0.0025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructor</td>
<td>0.0011</td>
<td>207.2063</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Professor Sex</td>
<td>Progress</td>
<td>0.3739</td>
<td>152.7634</td>
<td>0.0024</td>
<td>0.0306</td>
</tr>
<tr>
<td></td>
<td>Course</td>
<td>11.7156</td>
<td>322.0000</td>
<td>0.0364</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructor</td>
<td>19.2889</td>
<td>364.3072</td>
<td>0.0529</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 28, professor gender-role orientation explains 18.11% of the variability of student ratings of professor competence, student sex explains 0.40% of the variability of student ratings of professor competence, and professor sex explains 3.06% of the variability of student ratings of professor competence.
Because of the significant difference in these coefficients of determination, the null hypothesis is rejected and the research hypothesis is supported. Stated another way, it appears that professor gender-role orientation explains a much greater proportion of variability in the student ratings of professor competence than do the sex of the student or the sex of the professor.

Summary of Results

This study yielded a tremendous amount of data and output. Even when presented question by question, the overall findings and meaning of the data analyses may not be readily apparent or easily recalled. Table 29 is therefore included as a summary of the results. It presents each original research question, a brief summary of the findings, and an answer to the research question.

In making sense of these findings, it was essential to address the relevance of research design as well as contextual issues such as sociopolitical issues. In chapter V, an exploration of the meanings and potential implications will therefore be addressed.
<table>
<thead>
<tr>
<th>Research Question</th>
<th>Significance Found</th>
<th>Pattern of Findings</th>
<th>Answer to Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Is there a significant difference between ratings of professor competence by male and female students?</td>
<td>2 Factor ANOVA: 1 of 6 significant 1 Factor ANOVA: 0 of 3 significant</td>
<td>Female &gt; Male</td>
<td>NO</td>
</tr>
<tr>
<td>2 Is there a significant difference between ratings received by male and female professors?</td>
<td>2 Factor ANOVA: 1 of 6 significant 1 Factor ANOVA: 2 of 3 significant</td>
<td>Male &gt; Female</td>
<td>?</td>
</tr>
<tr>
<td>3 Is there a significant difference among the ratings received by professors who are perceived as masculine, feminine, androgynous, or undifferentiated by students?</td>
<td>2 Factor ANOVA: 6 of 6 significant 1 Factor ANOVA: 3 of 3 significant</td>
<td>A&gt;F&gt;M&gt;U</td>
<td>YES</td>
</tr>
<tr>
<td>4 Are there interaction effects between student and professor sex, between professor sex and gender-role orientation, or between student sex and professor gender-role orientation?</td>
<td>No Model significance when including interaction variables in model</td>
<td></td>
<td>NO</td>
</tr>
<tr>
<td>5 Does sex or gender-role orientation explain a greater proportion of the variability in the student ratings of professor competence?</td>
<td>GRO &gt; prof sex &gt; student sex</td>
<td>GRO explains the greatest proportion</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER V

DISCUSSION OF FINDINGS: IMPLICATIONS AND EXTENSIONS

Introduction

This chapter will begin by briefly revisiting the initial purpose and design of the study. It will then turn to an interpretation and discussion of the findings as stated in Chapter IV. Each set of findings will be presented with the research question being addressed, the research hypothesis at the start of the study and the results. A discussion of each set of findings with respect to the research question and hypothesis will then be provided. Specifically, this discussion will involve speculation about the possible meaning of each set of findings and the implications of those findings when answering the initial research questions.

Because there will be uncertainty and ambiguity with regard to the implications of the findings, this chapter will then provide a discussion of limitations of this current study as well as suggestions for future research. A summary of this researcher's conclusions will close this chapter.

Purpose and Design of the Study

This study was designed for the purpose of determining whether student evaluations of professors continue to be biased by sex and/or gender-role orientation at
the present time. In exploring these issues, this study also sought to provide information regarding the existence and operation of these dynamics at the graduate school level and specifically targeted one particular graduate department of Counselor Education and Counseling Psychology. Unique features of this study consisted of the inclusion of both sex and gender-role orientation as independent variables, of the choice to measure professor gender-role orientation using student perceptions rather than professor self-report, of the exploration of the research questions at the graduate school level, of the focus on a field which has purported to be a leader with regard to gender and other diversity issues, of the use of a unique student evaluation instrument which allowed for three separate measures of student ratings, and of the choice to conduct a study specific to and useable by one particular department.

A sample of 249 male and female graduate students completed a set of three instruments: (1) the Instructional Development and Effectiveness Assessment (IDEA) System to evaluate their professor, (2) a modified version of the Bern Sex-Role Inventory (BSRI) Short Form to describe their professor, and (3) a questionnaire designed by the researcher to provide additional data. This data consisted of demographic data allowing each participant to be categorized as a male or female for use in later analysis and one item in which the student rated the overall quality of the course.

Professors completed two instruments: (1) the IDEA Faculty Information Form (an instrument to allow them to assign weights to reflect the importance of 10 student progress items on the IDEA instrument), and (2) the BSRI Short Form to describe themselves. Each professor was categorized in three ways for later analysis: (1) as a
The data were then analyzed using analyses of variance (ANOVARs) to identify any main effects and/or interaction effects of student sex, professor sex, and professor gender-role orientation as rated by the students on each of the three measures of professor competence collected from the students. In the initial design, the use of three separate 2x2x4 ANOVAs was anticipated (one for each of the three dependent variables). However, as discussed in Chapter III, it was necessary to instead utilize three separate sets of 2-Factor ANOVAs in order to achieve model significance and to yield useable results.

It is to a discussion and interpretation of these results that this chapter now turns.

**Interpretation and Discussion of Results**

**Results Relevant to Research Question 1**

It will be recalled that the first research question addressed in this study was: "Is there a significant difference between ratings of professor competence provided by male and female students?" In essence, answering this question involved determining whether a main effect of student sex was evidenced by the data. The research hypothesis represented the researcher's initial belief prior to the collection and analysis of data and stated that "there will be no significant difference between ratings of professor
competence by male and female students." In other words, no main effect of student sex was anticipated by the researcher.

Summary of Results

Testing for a main effect involved running three 2x2 ANOVAs in which student sex and professor sex were included in the model and running three 2x4 ANOVAs in which student sex and professor gender-role orientation were included in the model. In the first set, no model significance was found. This indicates that no main effects of student sex were found in the set of 2x2 ANOVAs. In the second set of ANOVAs, model significance was found and the results from the three ANOVAs indicated a main effect on only one ANOVA. Specifically, a main effect of student sex was found only when both student sex and professor gender-role orientation were included in the model and when student self-reported progress was utilized as the dependent variable.

In interpreting the implications of these results for answering Research Question 1, it is interesting to note that, in all cases (all six ANOVAs), female students tended to give higher ratings than did male students. However, this pattern was not statistically significant and may be attributable to chance differences in ratings. In only one case (as described above) was a statistically significant difference found.

The 1-Factor ANOVAs on student sex yielded no model significance and no indication of a main effect of student sex. Therefore, the results of this research study were consistent with the research hypothesis and failed to support the existence of a main effect of student sex.
Discussion

This suggests that, although slight differences may exist in the ratings provided by male and female students, there was no evidence to suggest a consistent sex bias with regard to student sex. Male and female students in the department studied may therefore be viewed as offering similar ratings of professors.

Results Relevant to Research Question 2

It will be recalled that the second research question addressed in this study was: “Is there a significant difference between ratings of professor competence received by male professors and female professors?” In essence, answering this question involved determining whether a main effect of professor sex was evidenced by the data. The research hypothesis represented the researcher’s initial belief prior to the collection and analysis of data and stated that “male professors will receive higher ratings of competence from students than will female professors.” In other words, a main effect of professor sex was anticipated by the researcher.

Summary of Results

Testing for a main effect involved running three 2x2 ANOVAs in which professor sex and student sex were included in the model and three 2x4 ANOVAs in which professor sex and professor gender-role orientation were included in the model. In the first set of ANOVAs, no model significance was found. This indicates that no main
effects of student sex were found in the set of 2x2 ANOVAs. In the second set of 2-Factor ANOVAs, model significance was found and the results from these three ANOVAs indicated that a main effect of professor sex existed only when the instructor rating was used as the dependent variable. In this particular case, instructor ratings received by male professors were statistically significantly higher than those received by female instructors.

As described in Chapter IV, two main effects (using course rating and instructor rating) were found using a set of 1 Factor ANOVAs. When using course rating and instructor rating as dependent variables, males received higher ratings than did females.

To summarize the results, therefore, it appears that only one of six 2-Factor ANOVAs yielded a main effect of professor sex and that two of three 1-Factor ANOVAs yielded a main effect of professor sex. In each case in which a main effect was found, males received higher mean ratings than did females.

Discussion

The implications of these results to Research Question 2 remain ambiguous. In interpreting these results, it is helpful to consider two avenues of thinking: (1) the impact of including gender-role orientation in the model, and (2) and the impact of using student self-reported progress as a dependent variable.

When including professor gender-role orientation in the model, there does not appear to be enough evidence to support the consistent existence of a main effect of professor sex. This makes sense in light of the researcher’s fifth research hypothesis.
which posited that gender-role orientation would explain a greater proportion of variability in student ratings than would the sex of either the professor or the student. What this suggests is that past research on sex bias may have found inconsistent evidence of sex bias because of its use only of sex as an independent variable. Because gender-role orientation appears to explain a far greater proportion of the variability in student ratings, the inclusion of this variable in the model overshadows the importance of sex as a potentially biasing factor. Given the overall findings of this study, the researcher suggests that the results do not support a conclusion that professor sex is instrumental in terms of explaining the variability in student ratings of professor competence. Instead, the researcher concludes that the answer to Research Question 2 is that there is not a significant difference between the ratings received by male and female professors.

In addition to supporting this conclusion, the data further suggest that one way in which the academic department might limit the impact of any existing sex bias would be to utilize as at least one measure of professor effectiveness the student self-reports of progress with respect to a variety of learning objectives. Consistently, no main effects of professor sex were noted when the progress rating was utilized as the dependent variable. This suggests the possibility that the progress ratings measure may be less affected by professor sex than are the instructor ratings and the course ratings.

Results Relevant to Research Question 3

It will be recalled that the third research question addressed in this study was: "Is there a significant difference among the ratings received by professors who are
perceived as masculine, feminine, androgynous, or undifferentiated by students?" In essence, answering this question involved determining whether a main effect of professor gender-role orientation as perceived by students was evidenced by the data. The research hypothesis represented the researcher's initial belief prior to the collection and analysis of data and stated that there would be a significant difference and that, specifically, “professors who are perceived as masculine or androgynous will receive higher ratings from students than those who are perceived as feminine or undifferentiated.” In other words, a main effect of professor gender-role orientation was anticipated by the researcher and a hypothesis about the order of ratings was specified.

**Summary of Results**

Testing for main effects of professor gender-role orientation involved running a set of three 2x4 ANOVAs in which professor gender-role orientation and student sex were included in the model and a set of three 2x4 ANOVAs in which professor gender-role orientation and professor sex were included in the model. In each of these six ANOVAs, there was evidence for a main effect of professor gender-role orientation at the \( p < 0.0001 \) level. In addition, each of the 1-Factor ANOVAs yielded \( p < 0.0001 \) values for main effects of professor gender-role orientation.

The order in which ratings were received according to gender-role orientation was also consistent across all nine analyses. Specifically, professors who were perceived as androgynous received the highest ratings, followed (in order) by those perceived as feminine, masculine, and undifferentiated.
Discussion

These results clearly indicate that there is a main effect of professor gender-role orientation on student perceptions of professor competence. When combined with the results relevant to Research Question 5, it is apparent that professor gender-role orientation as perceived by students explains a significant proportion of variability in the ratings of professor competence. In effect, because gender-role orientation is a characteristic considered unrelated to objective standards of teaching effectiveness, these results are therefore indicative of a bias.

Definitive results indicative of a bias based on gender-role orientation were, indeed, anticipated by the researcher. What was not anticipated was the order of preference with regard to gender-role orientation. The researcher had anticipated that professors perceived as androgynous or masculine would be the most highly rated; instead those perceived as androgynous or feminine actually received the highest ratings. This unexpected finding warrants additional discussion.

The researcher’s initial thinking was that masculine and androgynous gender-role orientations would receive the highest ratings and be most associated with male professors. Consistent with this expectation, nearly 38 percent of male professors were viewed as androgynous whereas only 15 percent of female professors were viewed as androgynous. However, contrary to the researcher’s expectations, more female professors were perceived as masculine than were male professors. In actuality, 28 percent of the female professors were perceived as masculine whereas only 22 percent
of the male professors were perceived as masculine. In addition, 30 percent of the female faculty were perceived as undifferentiated.

One possible explanation for these surprising results is that a confounding variable might be chronological age. Data regarding the actual age of professors were not collected as part of this study. However, within the department in which this research was conducted, the male professors are noticeably older than the female professors. Developmental psychology has suggested that adherence to gender-role expectations (sex-typed behavior) decreases with aging in the adult lifespan. Therefore, the tendency for male professors in this study to be perceived as androgynous more often than masculine might be reflective of actual changes related to aging. However, this would not contribute an understanding of the females' tendency to be perceived as undifferentiated, masculine or feminine.

These findings are interesting in light of the contentions posed by Basow and Silburg (1987) and by Martin (1984). These researchers suggested that, because professors may be expected to behave in stereotypic ways, a female professor may be expected to be warm and nurturant whereas a male professor may not. Should the warmth and nurturance demonstrated by both a male and female be approximately equal, the male's behavior may then be more noticeable. Similarly, should the competence, autonomy, and competitiveness demonstrated by both a male and female be approximately equal, the female's behavior may then be more noticeable.

Although speculative, this concept of judging actual behavior and characteristics in terms of gender stereotypic expectations may be useful in understanding the
unexpected finding that more males were perceived as androgynous and more females were perceived as masculine. It also revisits the question initially posed by Morrison, et al. (1992): how do female faculty members successfully balance expectations and fall within a very “narrow band of acceptable behaviors” (p. 47)? (See Appendix B.) It appears that female faculty members within the department studied may have been viewed by students as better adjusting to the expectation of stereotypically masculine behaviors and characteristics and have been perceived as not demonstrating enough stereotypically feminine characteristics. In the sample studied, more female professors tended to be low in feminine characteristics as evidenced by them being perceived as either masculine (high masculine, low feminine) or undifferentiated (low masculine, low feminine) than to be high in feminine characteristics. Clearly, female professors were underrepresented in the androgynous category, suggesting that it is most challenging for female professors to be perceived as demonstrating both a high level of masculine and a high level of feminine characteristics.

Any such speculation, however, needs to remain just that. Considering the fact that no significant interaction effects between sex and gender-role orientation were found, the trend of males being perceived as androgynous and females being viewed as undifferentiated or masculine was not statistically significant in this sample. Furthermore, although there remains ambiguity with regard to Research Question 2, the researcher concluded that there was sufficient evidence to suggest a main effect professor sex.

To summarize this discussion, professor gender-role orientation appears to significantly influence student perceptions of professor effectiveness and should be
considered a potential source of bias when interpreting student evaluations. Regardless of the dependent variable use, those professors who were perceived as androgynous (primarily male professors) and feminine (fairly even split between male and female professors) were rated significantly higher than those professors who were viewed as masculine or undifferentiated (primarily female professors).

It may be useful for the department or individual professors to consider both the utility of student evaluations given this apparent bias and the implications in terms of self-presentation. In addition, from a sociological perspective, an exploration of how people are socialized to have narrow sex-role expectations would be warranted. Finally, subsequent analysis of the data produced might be useful in examining exactly which characteristics were attributed to male and female professors and how individual characteristics contained within the BSRI relate to the actual rating of professors.

Results Relevant to Research Question 4

It will be recalled that the fourth research question addressed in this study was: “Are there interaction effects between student sex and professor sex, between professor sex and gender-role orientation, or between student sex and professor gender-role orientation?” In essence, answering this question involved determining whether any interaction combinations yield a significant effect. The research hypothesis represented the researcher’s initial belief prior to the collection and analysis of data and stated that “interaction effects will exist....” Specifically, the researcher hypothesized that male professors would tend to be seen as masculine or androgynous, that female professors
would tend to be seen as androgynous or feminine, that male professors would receive higher ratings from male students, and that female professors would receive higher ratings from female students.

Summary of Results

Because neither the 3-Factor nor the 2-Factor ANOVA models were significant when the interaction variables were included in the models, it was concluded that there were no significant interaction effects. This finding was contrary to the research hypothesis.

Discussion

Based on these findings, it can be concluded that no significant interaction effects existed in the sample studied. This was helpful in allowing for less ambiguous understanding of the tests for main effects. Although there may be some relationship between professor sex and professor gender-role orientation as speculated in the discussion of the results relevant to Research Question 3, it appears that the relationship is weak.

Results Relevant to Research Question 5

It will be recalled that the fifth and final research question addressed in this study was: “Does sex or gender-role orientation explain a greater proportion of the variability in the student ratings of professor competence?” In essence, answering this question
involved computing and comparing the coefficients of determination for student sex, professor sex, and professor gender-role orientation. The research hypothesis represented the researcher’s initial belief prior to the collection and analysis of data and stated that “Gender-role orientation will explain a greater proportion of the variability in the student ratings of professor competence than will the sex of the professor or student.”

**Summary of Results**

Indeed, as shown in Table 28, gender-role orientation was found in this study to explain a significantly greater proportion of variability in student ratings of professor competence than did professor sex or student sex. Professor gender-role orientation explained, on average across the three dependent variables, 18.11 percent of the variability; professor sex explained, on average, 3.06 percent; and student sex, on average, explained only 0.4 percent of the variability of student ratings.

**Discussion**

This finding, in combination with the findings relevant to Research Question 3, is of great import to the conclusions of this study. It appears that gender-role orientation, rather than biological sex, is much more helpful in understanding and explaining different ratings received by professors in the department studied. It suggests that the characteristics measured by the BSRI, which are intended to be at least stereotypically gender related, do indeed relate to a student’s perceptions of a professor’s
competence.

How the gender-role orientation relates to or influences student perceptions, however, remains unclear. As discussed earlier, it may be that women face a particular challenge in terms of meeting a narrow band of expectations. It may be that professors in Counselor Education and Counseling Psychology are expected to demonstrate a high number of both masculine and feminine characteristics and that these expectations are somewhat unique to the field. It may be that age interacts with gender-role orientation.

The possibilities are numerous. Amidst these possibilities, though, there is clearly an indication that gender-role orientation appears to have a significant impact on student perceptions of professor competence in the department studied. What the dynamics are which underlie such an impact remain unclear and suggest the need for future research. Before addressing specific recommendations for future research, however, it is necessary to first address some of the limitations of the current study.

Limitations

As stated earlier in this dissertation, there were a number of limitations to this study. Some of these limitations were due to conscious decisions to delimit the scope of the study in terms of variables. Other limitations were related to decisions regarding breadth of data collection. Still other limitations were due to choice of instrumentation. At this point, it is helpful to revisit these limitations, to speculate about their potential relationship to the findings, and to look toward future research.

The first limitation of this study was due to a conscious decision to delimit the
scope of this study with regard to variables studied. By examining the impact only of sex and gender-role orientation on student perceptions of professor competence, the potential of yielding results which would thoroughly investigate possible influences on student evaluations was necessarily limited. Other variables such as age, height, weight, years of teaching experience, nature and extent of a professor's other responsibilities (e.g., family), or status with regard to tenure were not included in this study.

A second limitation of this study was also related to the breadth and scope of the study. In deciding to study only teaching evaluations (which represent only one area upon which ratings of overall professor competence are ultimately based) and to utilize only student evaluations in this examination of teaching evaluations was clearly a limitation. The impact of sex and gender-role orientation on the evaluations of a professor's teaching competence by individuals other than students is unknown and largely unresearched. Furthermore, the impact of sex and gender-role orientation on overall evaluations of a professor's competence (research, teaching, service) is also unknown and largely unresearched.

A third limitation of this study is its limited generalizability. This research was conducted in one department at one university in order to have results which were very specific and applicable to that department. Attempts to generalize the results of this study to other departments of Counselor Education or Counseling Psychology, to other graduate departments, or to other universities should be made with great caution until this research is replicated across a number of settings.

A fourth limitation of this study also related to generalizability has to do with the
applicability of this study for understanding the overall dynamics of the glass ceiling effect. As described earlier (pp. 10-11), the glass ceiling effect appears to involve a number of dynamics and to be quite complex. Indeed, there a number of theories which attempt to explain it from varying perspectives. In studying the impact of sex and gender-role orientation on student perceptions of professor competence, this research may contribute to a greater understanding of the sex-based discrimination by society at large. This research would therefore address, but not explain, glass ceiling effects from the perspective of the theory which contends that “discrimination by the majority population [is]... the major cause of the inequities” (Morrison & von Glinow, 1990. p. 201).

A fifth limitation is related to the choice of instrumentation. The IDEA rating form was chosen after great consideration as the primary measure of student ratings of professor competence. This instrument provides three global measures of professor competence: a self-report of progress or learning, a rating of the instructor, and a rating of the course. It is a well-researched, validated, and reliable instrument which is widely used across North America. Nonetheless, the selection of this particular instrument may have had an influence on the results of this research. How sex and gender-role orientation would impact student ratings of professors using the department’s end-of-course survey or another published instrument cannot be determined using the data collected in this study. Therefore, generalizations regarding the impact of these variables on ratings using other instruments should be made with great caution. Furthermore, it will be recalled that the researcher chose to create a separate item (Question 45 on the
Student Questionnaire) to measure the student rating of the quality of the course rather than using the item on the IDEA form. The impact of this decision on the results of this study are also unknown. Therefore, conclusions regarding the impact of sex and gender-role orientation on student ratings of course quality should be made with this fact in mind.

These five limitations represent the most apparent limitations of the study but are not intended to provide exhaustive coverage of all possible limitations. Rather, they are intended to serve as the basis from which this chapter now provides suggestions for future research.

Suggestions for Future Research

A review of the findings and limitations of this research study would suggest a need for the expansion as well as replication of this research. Based on the findings and limitations of the current study, the researcher would recommend that similar but more comprehensive studies be undertaken both at the department level (in institutions desiring results specific to their setting) and at a broader level involving numerous universities. Collection of additional data on professors such as their age, tenure status, years of experience, and nature/extent of their other responsibilities would allow for more complicated analyses of factors which may influence student perceptions of their competence. Such research might also utilize instrumentation which would provide additional data on the professor's personality style, teaching style, or leadership style.
Conducting such studies in other settings would allow for the replication of this study, the testing of the generalizability of these results, and the inclusion of a wide variety of variables not included in this study.

Another recommendation for future research would be to enhance the depth of information collected from students. It would be interesting to combine both quantitative methods (as used in this study) with qualitative methods such as focus groups or structured interviews in order to better understand the experiences of students and the thoughts or feelings contributing to the rating of a specific professor.

As implied in the discussion of limitations, this researcher would also recommend the use of a variety of end-of-course surveys to better understand the impact of sex and gender-role orientation on various rating systems. She would also suggest the use of instruments designed to assess how other faculty, chairpersons, or administrators (all of whose opinions have considerable impact on the tenure and promotion process) rate a professor and how sex and gender-role orientation may impact these ratings. These individuals are in a position to rate not only teaching effectiveness, but also scholarly productivity and service performance.

A final suggestion for future research is to focus on faculty experiences. Obtaining information, both quantitatively and qualitatively, from individual professors regarding their valuing (or lack thereof) of student perceptions, their teaching styles and behaviors, their actual ratings by students, and their reactions to these ratings would be both a valuable research endeavor as well as a useful technique to transform otherwise summative evaluations into a more formative evaluation process.
Summary and Conclusions

This study examined the impact of sex and gender-role orientation on student evaluations of teaching effectiveness in a Counselor Education and Counseling Psychology department at one large university in the Midwest. The findings provide strong support for the hypothesis that professor gender-role orientation as perceived by students has a statistically significant impact on student ratings of the quality of the instructor, the quality of the course, and of their self-reported progress in the course. The findings also provide strong support for the hypothesis that gender-role orientation explains a significantly greater proportion of the variability in student ratings than do either professor sex or student sex.

While these findings may be used to question the validity or utility of student evaluations, they may also be used to facilitate a deeper understanding of student perceptions and expectations or to serve a more formative function in which a professor might choose to alter his or her behaviors. The decision of how to interpret and respond to these findings is recognized as a deeply personal and individual one and as having definite sociopolitical implications.
Appendix A

Rules for Teachers, 1872
Rules for Teachers, 1872

1. Teachers each day will fill lamps and clean chimneys.
2. Each teacher will bring a bucket of water and scuttle of coal for the day's session.
3. Make your pens carefully. You may whittle nibs to the individual taste of the pupils.
4. Men teachers may take one evening each week for courting purposes, or two evenings a week if they go to church regularly.
5. After 10 hours of school, the teacher may spend the remaining time reading the Bible or other good books.
6. Women teachers who marry or engage in unseemingly conduct will be dismissed.
7. Every teacher should lay aside from each pay a goodly sum of his earnings for his benefit during his declining years so that he will not become a burden on society.
8. Any teacher who smokes, uses liquor in any form, or frequents pool or public halls or gets shaved in a barber shop will give good reason to suspect his worth, intentions, integrity, and honesty.
9. The teacher who performs his labor faithfully and without fault for five years will be given an increase of 25 cents per week in his pay, providing that the Board of Education approves.
Appendix B

A Narrow Band of Acceptable Behaviors for Women
Masculine or Like Men

Feminine or Unique to Women

Acceptable Band

Appendix C

Letter of Permission to Utilize the Instructional Development and Effectiveness Assessment (IDEA) Instrument
February 14, 1995

Suzanne Hobson
1070 Claymoor Dr. 2-D
Kalamazoo, MI 49009

Dear Ms. Hobson:

This is to respond to your letter of February 6, 1995. On behalf of this Center, you and Western Michigan University are given permission to modify and use IDEA items copyrighted by this Center. This permission is given under the following conditions:

1. that they be used only for research related to your doctoral dissertation.

2. that the Center for Faculty Evaluation and Development and Kansas State University be indicated as their source and that the Center's copyright be acknowledged. (I suggest that your note read: "Copyrighted by the Center for Faculty Evaluation and Development, Kansas State University. Adapted with permission.")

3. that a copy of this letter be included in your dissertation.

4. that the Center will receive one complimentary copy of your dissertation when completed—an unbound copy is acceptable.

Sincerely yours,

William E. Cashin, Ph.D.
Director

center for
EVALUATION & DEVELOPMENT
1615 Anderson Avenue, Manhattan, KS 66502-4073
Toll-Free 800-255-2757
or 913-532-5970
Division of Continuing Education
Kansas State University

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix D

HSIRB Letter of Approval
Date: April 10, 1995
To: Hobson, Suzanne M.
From: Richard Wright, Interim Chair
Re: HSIRB Project Number 95-04-10

This letter will serve as confirmation that your research project entitled “The impact of sex and gender-role orientation on student evaluations of professor competence in counselor education and counseling psychology” has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you must seek specific approval for any changes in this design. You must also seek reapproval if the project extends beyond the termination date. In addition if there are any unanticipated adverse or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: Apr 10, 1996

xc: Trembley, Edward, CECX
Appendix E

Letter of Permission to Utilize Bem Sex-Role Inventory (BSRI) Instrument
Bem Inventory
Test Booklet (Short and Original)
Permission to reproduce for one year starting from date of purchase: February 14, 1995

by Sandra Lipsitz Bem

Distributed by Mind Garden
P.O. Box 60669 Palo Alto California 94306 (415) 424-8493

Copyright © 1978 Consulting Psychologists Press, Inc. All rights reserved.

It is your legal responsibility to compensate the copyright holder of this work for any reproduction in any medium. The copyright holder has agreed to grant one person permission to reproduce this work for one year (a maximum of 200 copies) from the date of purchase for non-commercial and personal use only. Non-commercial use means that you will not receive payment for distributing this document and personal use means that you will only reproduce this work for your own research or for clients. This permission is granted to one person only. Each person who administers the test must purchase permission separately. Any organization purchasing permissions must purchase separate permissions for each individual who will be using or administering the test.
Appendix F

Student Questionnaire
STUDENT QUESTIONNAIRE

Please respond to each of the following 5 items using the code which accompanies each item.

45. On a scale of 0-4 (with 0 meaning the lowest quality and 4 meaning the highest quality), I rate the quality of this course a ______.

46. I am in the ___ program:
   0 = Master's in Community Agency Counseling
   1 = Master's in Student Personnel Services in Higher Education
   2 = Master's in Counselor Education and Supervision
   3 = Master's in Pupil Personnel Services (School Counseling)
   4 = Master's in Counseling Psychology
   5 = Master's in CECP - Specialty unknown or undecided
   6 = Doctoral in Counselor Education/Community Agency
   7 = Doctoral in Counseling Psychology
   8 = None of the Above

47. I consider myself a _______ student.
   0 = Full-time
   1 = Part-time

48. My current employment status is: ______.
   0 = Full-time
   1 = Part-time
   2 = Not employed

49. My sex is: ____.
   0 = Female
   1 = Male

50. The sex of the instructor teaching this course is: ____.
   0 = Female
   1 = Male

THANK YOU FOR YOUR TIME!
Appendix G

Oral Recruitment Script
Oral Recruitment Script

Hello. My name is Suzanne Hobson and I'm a doctoral student in the Counseling Psychology Program. I'd like to thank (insert professor's name) for allowing me to visit your class today. I'm here today to ask each of you to participate in my dissertation research. If you agree to participate, you will be asked to complete three questionnaires containing a total of 50 questions. This should take approximately 20 minutes and I will be having you complete them here in class today. If you have already participated in my study, you are not eligible to participate again.

The purpose of my dissertation research is to investigate student perceptions of instructor competence. Specifically, I am interested in the possible relationship between a number of student and professor characteristics and styles and student evaluations of professor competence. I am interested in this research topic because I plan to become a professor after I graduate and I think it's really important to consider what students view as effective and not effective. This is an opportunity for you to contribute to the scientific knowledge base regarding effective teaching.

You need to know that your participation is completely voluntary. There will be no penalty if you choose not to participate. You also need to know that your responses will be completely anonymous. Your responses will be identified only by an arbitrary code, and I will have no way of knowing who completed which response sheet. Furthermore, there will be no way for me to know how this class as a whole rated your professor. Therefore, no information from my research will be available to your professor or to the CECP department about how your specific professor was rated by this class.

At this time, I will distribute materials to each of you. I will give you a set of three items: (1) an Informed Consent Information Form for you to read; (2) the set of questionnaires; and (3) a Scantron Answer sheet. If, after reading the Informed Consent Form, you are willing to participate, you may proceed by completing the three questionnaires. Please do not put your name on the answer sheet or on the questionnaire. Also, please complete the questions in order. Do not look ahead or complete the next instrument until you have completed all items preceding it. If you are not participating, you are asked to simply sit quietly while your classmates participate. No one will be allowed to leave during this time.
Appendix H

Informed Consent Form
INFORMED CONSENT INFORMATION FORM

Principal Investigator: Edward Trembley, D.Ed.  
Research Associate: Suzanne Hobson, M.A.

I have been invited to participate in a research project investigating student perceptions of professor competence in the Counselor Education and Counseling Psychology fields. I understand that the purpose of this study is to examine some of the factors which may influence student perceptions of professor competence. I further understand that this project is intended to fulfill Suzanne Hobson's dissertation requirement.

My consent to participate in this project indicates that I will be asked to complete a series of three questionnaires: the IDEA short form, the Bem Inventory short form, and a demographic questionnaire during today's class period.

As in all research there may be unforeseen risks to the participant. If an accidental injury occurs, appropriate emergency measures will be taken; however, no compensation or treatment will be made available to me except as otherwise specified in this consent form. I understand that the only anticipated risk to me involves the inconvenience of completing the questionnaires and that this is expected to require approximately 20 minutes.

One way in which I may benefit from this activity is having the opportunity to contribute to research which may enhance professor understanding of what students perceive as effective instruction. Also, I may gain a feeling of satisfaction from helping a student completing her dissertation research.

I understand that all the information collected from me is anonymous. That means that my name will not appear on any papers on which this information is recorded. The forms will not be coded and there will no way to identify my responses. All forms will be retained for three years in a locked file in the principal investigator's laboratory.

I understand that I may refuse to participate or quit at any time during the study without prejudice or penalty. If I have any questions or concerns about this study, I may contact either Suzanne Hobson at 372-9768 or Dr. Trembley at 387-5115. I may also contact the Chair of Human Subjects Institutional Review Board or the Vice President for Research if questions or problems arise during the course of the study. My completion of the questionnaires indicates that I understand the purpose and requirements of the study and that I agree to participate.
Appendix I

Results of 2 Factor ANOVAs
Data Analysis Output from SAS Program

Two Factor ANOVA including Professor Sex and Professor Gender-role Orientation in Model

General Linear Models Procedure: Class Level Information

Class Levels Values
GENDER_P 2 Female Male
GRO_BYS 4 Androgynous Feminine Masculine Undifferentiated

Number of observations in data set = 239

Test of Model Significance: Progress Rating as Dependent Variable

The SAS System 2

General Linear Models Procedure

Dependent Variable: PROGRESS

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>26.31855739</td>
<td>12.48</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>234</td>
<td>123.38422138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>238</td>
<td>149.70277878</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square C.V. PROGRESS Mean
0.175805 26.81081 2.70839444

ANOVA Summary Table: Progress Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>0.36829887</td>
<td>0.70</td>
<td>0.4041</td>
</tr>
<tr>
<td>GRO_BYS</td>
<td>3</td>
<td>26.21062335</td>
<td>16.57</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
### Marginal Means: Progress Rating as Dependent Variable

#### Least Squares Means

| GENDER_P | PROGRESS | Pr > |T| H0: LSMEAN1=LSMEAN2 |
|----------|----------|------|----------------------|
| Female   | 2.75781158 | 0.4041 |
| Male     | 2.57737672 |

#### GRO_BYS

| GRO_BYS            | PROGRESS | Pr > |T| H0: LSMEAN(i)=LSMEAN(j) |
|--------------------|----------|------|-------------------------|
|                    |          | i/j  | 1 | 2 | 3 | 4 |
| Androgynous        | 3.16782284 | 1 | . | 0.0062 | 0.0007 | 0.0001 |
| Feminine           | 2.78894351 | 2 | 0.0062 | . | 0.5122 | 0.0001 |
| Masculine          | 2.70295830 | 3 | 0.0007 | 0.5122 | . | 0.0002 |
| Undifferentiated   | 2.21065194 | 4 | 0.0001 | 0.0001 | 0.0002 | . |

**NOTE:** To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
Test of Model Significance: Course Rating as Dependent Variable

General Linear Models Procedure

Dependent Variable: COURSE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>49.14922471</td>
<td>10.87</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>234</td>
<td>264.58299286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>238</td>
<td>313.73221757</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.156660
C.V.: 35.84468
COURSE Mean: 2.96652720

ANOVA Summary Table: Course Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>2.95757440</td>
<td>2.62</td>
<td>0.1072</td>
</tr>
<tr>
<td>GRO_BYS</td>
<td>3</td>
<td>40.45363537</td>
<td>11.93</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Marginal Means: Course Rating as Dependent Variable

| GENDER_P  | COURSE | Pr > |T| H0: LSMEAN1=LSMEAN2 |
|-----------|--------|------|---------------------|
| Female    | 2.87604590 | 0.1072 |
| Male      | 3.10398140 |

| GRO_BYS     | COURSE | Pr > |T| H0: LSMEAN(i)=LSMEAN(j) |
|-------------|--------|------|-------------------------|
| Androgynous | 3.56477360 | 1 . 0.0394 0.0003 0.0001 |
| Feminine    | 3.14852903 | 2 0.0394 . 0.1110 0.0002 |
| Masculine   | 2.84167793 | 3 0.0003 0.1110 . 0.0233 |
| Undifferentiated | 2.40507404 | 4 0.0001 0.0002 0.0233 . |

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Test of Model Significance: Instructor Rating as Dependent Variable

IThe SAS System

General Linear Models Procedure

Dependent Variable: TEACHER

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>71.41868921</td>
<td>14.69</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>234</td>
<td>284.39930242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>238</td>
<td>355.81799163</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.200717, C.V.: 37.72140, TEACHER Mean: 2.92259414

ANOVA Summary Table: Instructor Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>6.57271784</td>
<td>5.41</td>
<td>0.0209</td>
</tr>
<tr>
<td>GRO_BYS</td>
<td>3</td>
<td>56.12504024</td>
<td>15.39</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Marginal Means: Instructor Rating as Dependent Variable

General Linear Models Procedure

Least Squares Means

<table>
<thead>
<tr>
<th>GENDER_P</th>
<th>TEACHER</th>
<th>Pr &gt;</th>
<th>T</th>
<th>H0:</th>
<th>LSMEAN</th>
<th>LSMEAN1=LSMEAN2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>2.78313125</td>
<td>0.0209</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3.12292598</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| GRO_BYS | TEACHER | Pr > |T| H0: LSMEAN(i)=LSMEAN(j) | LSMEAN(i) | LSMEAN(j) |
|---------|---------|------|--------------------------|-----------|-----------|
| Androgynous | 3.57475900 | 1 . | 0.1598 | 0.0001 | 0.0001 |
| Feminine   | 3.28098632 | 2 0.1598 | . | 0.0010 | 0.0001 |
| Masculine  | 2.61950915 | 3 0.0001 | 0.0010 | . | 0.1551 |
| Undifferentiated | 2.33686001 | 4 0.0001 | 0.0001 | 0.1551 | . |

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
Two Factor ANOVA including Student Sex and Professor Gender-role Orientation in Model

General Linear Models Procedure: Class Level Information

1. The SAS System

General Linear Models Procedure
Class Level Information

Class Levels Values
GENDER_S 2 Female Male
GRO_BYS 4 Androgynous Feminine Masculine Undifferentiated

Number of observations in data set = 126

Test of Model Significance: Progress Rating as Dependent Variable

1. The SAS System

General Linear Models Procedure

Dependent Variable: PROGRESS

Source | DF | Sum of Squares | F Value | Pr > F
-------|----|----------------|---------|---------
Model 4 18.95231235 9.55 0.0001
Error 121 60.03594290
Corrected Total 125 78.98825524

R-Square C.V. PROGRESS Mean
0.239938 26.55342 2.65272659

ANOVA Summary Table: Progress Rating as Dependent Variable

Source | DF | Type III SS | F Value | Pr > F
-------|----|-------------|---------|---------
GENDER_S 1 2.29754931 4.63 0.0334
GRO_BYS 3 18.20141136 12.23 0.0001

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Marginal Means: Progress Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Least Squares Means

| GENDER | PROGRESS | Pr > |T| H0: LSMEAN1=LSMEAN2 |
|--------|----------|------|---------------------|
| Female | 2.80308110 | 0.0334 |
| Male   | 2.52863036 |      |

| GRO_BYS   | PROGRESS | Pr > |T| H0: LSMEAN(i)=LSMEAN(j) |
|-----------|----------|------|--------------------------|
| Androgynous | 3.20972728 | 1  | 0.0102 | 0.0004 | 0.0001 |
| Feminine | 2.74235907 | 2 | 0.0102 | . | 0.3089 | 0.0013 |
| Masculine | 2.55739918 | 3 | 0.0004 | 0.3089 | . | 0.0239 |
| Undifferentiated | 2.15393737 | 4 | 0.0001 | 0.0013 | 0.0239 | . |

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
Test of Model Significance: Course Rating as Dependent Variable

The SAS System
General Linear Models Procedure

Dependent Variable: COURSE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>41.73742058</td>
<td>9.46</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>121</td>
<td>133.46892863</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>125</td>
<td>175.20634921</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.238219
C.V.: 35.96004
COURSE Mean: 2.92063492

ANOVA Summary Table: Course Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_S</td>
<td>1</td>
<td>2.76873633</td>
<td>2.51</td>
<td>0.1157</td>
</tr>
<tr>
<td>GRO_BYS</td>
<td>3</td>
<td>41.30516699</td>
<td>12.48</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Marginal Means: Course Rating as Dependent Variable

| GENDER_S | COURSE | Pr > |T| H0: LSMEAN1=LSMEAN2 |
|----------|--------|------|---------------------|
| Female   | 3.08747084 | 0.1157 |        |
| Male     | 2.78618912  |        |        |

Least Squares Means

| GRO_BYS       | COURSE | Pr > |T| H0: LSMEAN(i)=LSMEAN(j) |
|---------------|--------|------|-------------------------|
| Androgynous   | 3.75641022 | 1 .  | 0.0149 0.0001 0.0001   |
| Feminine      | 3.09679484 | 2 0.0149 | 0.1490 0.0009 |
| Masculine     | 2.70481804 | 3 0.0001 | 0.1490 . 0.0522 |
| Undifferentiated | 2.18929684 | 4 0.0001 | 0.0009 0.0522 .|

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
Test of Model Significance: Instructor Rating as Dependent Variable

The SAS System 12

General Linear Models Procedure

Dependent Variable: TEACHER

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>41.60335196</td>
<td>7.60</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>121</td>
<td>165.60299725</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>125</td>
<td>207.20634921</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.200782; C.V.: 40.05568; TEACHER Mean: 2.92063492

ANOVA Summary Table: Instructor Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_S</td>
<td>1</td>
<td>1.20539260</td>
<td>0.88</td>
<td>0.3499</td>
</tr>
<tr>
<td>GRO_BTS</td>
<td>3</td>
<td>41.60228987</td>
<td>10.13</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Marginal Means: Instructor Rating as Dependent Variable

The SAS System 13

General Linear Models Procedure

Least Squares Means

GENDER_S         | TEACHER | Pr > |T| | HO: LSMEAN i=LSMEAN j |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>3.03530197</td>
<td>0.3499</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.83651118</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GRO_BTS | TEACHER | Pr > |T| | HO: LSMEAN(i)=LSMEAN(j) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Androgynous</td>
<td>3.66547385</td>
<td>1</td>
<td>0.2050</td>
<td>0.0005</td>
</tr>
<tr>
<td>Feminine</td>
<td>3.28654575</td>
<td>2</td>
<td>0.0262</td>
<td>0.0003</td>
</tr>
<tr>
<td>Masculine</td>
<td>2.60969692</td>
<td>3</td>
<td>0.0005</td>
<td>0.1467</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>2.18190979</td>
<td>4</td>
<td>0.0001</td>
<td>0.1467</td>
</tr>
</tbody>
</table>

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
Two Factor ANOVA including Professor Sex and
Student Sex in Model

General Linear Models Procedure: Class Level Information

1The SAS System

General Linear Models Procedure
Class Level Information

Class Levels Values
GENDER_S 2 Female Male
GENDER_P 2 Female Male

Number of observations in data set = 126

Test of Model Significance: Progress Rating as Dependent Variable

1The SAS System

General Linear Models Procedure
Dependent Variable: PROGRESS

Source DF Sum of Squares F Value Pr > F
Model 2 0.82330725 0.65 0.5250
Error 123 78.16494799
Corrected Total 125 78.98825524

R-Square C.V. PROGRESS Mean
0.010423 30.05114 2.65272659

ANOVA Summary Table: Progress Rating as Dependent Variable

Source DF Type III SS F Value Pr > F
GENDER_S 1 0.79755932 1.26 0.2648
GENDER_P 1 0.07240627 0.11 0.7363

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Marginal Means: Progress Rating as Dependent Variable

The SAS System

General Linear Models Procedure
Least Squares Means

| GENDER_P | PROGRESS | Pr > |T| H0: | LSMEAN | LSMEAN1=LSMEAN2 |
|----------|----------|------|------|--------|-----------------|
| Female   | 2.63724489 | 0.7363 |
| Male     | 2.68617587 |

| GENDER_S | PROGRESS | Pr > |T| H0: | LSMEAN | LSMEAN1=LSMEAN2 |
|----------|----------|------|------|--------|-----------------|
| Female   | 2.74204141 | 0.2648 |
| Male     | 2.58137935 |
# Test of Model Significance: Course Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Dependent Variable: COURSE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>3.51434780</td>
<td>1.26</td>
<td>0.2876</td>
</tr>
<tr>
<td>Error</td>
<td>123</td>
<td>171.69200141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>125</td>
<td>175.20634921</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.020058

C.V.: 40.45248

COURSE Mean: 2.92063492

## ANOVA Summary Table: Course Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_S</td>
<td>1</td>
<td>0.75410870</td>
<td>0.54</td>
<td>0.4637</td>
</tr>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>3.08209422</td>
<td>2.21</td>
<td>0.1399</td>
</tr>
</tbody>
</table>

## Marginal Means: Course Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Least Squares Means

| GENDER_P | COURSE | Pr > |T| H0: |LSMEAN1=LSMEAN2 |
|----------|--------|------|-----|----------------|
| Female   | 2.79131046 | 0.1399 |
| Male     | 3.11055159  |  |

| GENDER_S | COURSE | Pr > |T| H0: |LSMEAN1=LSMEAN2 |
|----------|--------|------|-----|----------------|
| Female   | 3.02904321 | 0.4637 |
| Male     | 2.87281883  |  |
Test of Model Significance: Instructor Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Dependent Variable: TEACHER

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>6.69549355</td>
<td>2.05</td>
<td>0.1326</td>
</tr>
<tr>
<td>Error</td>
<td>123</td>
<td>200.5108565</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>125</td>
<td>207.20634921</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square C.V. TEACHER Mean

0.032313 43.71587 2.92063492

ANOVA Summary Table: Instructor Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_S</td>
<td>1</td>
<td>0.12289984</td>
<td>0.08</td>
<td>0.7841</td>
</tr>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>6.69443147</td>
<td>4.11</td>
<td>0.0449</td>
</tr>
</tbody>
</table>

Marginal Means: Instructor Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Least Squares Means

| GENDER_P | TEACHER LSMEAN | Pr > |T| H0: LSMEAN1=LSMEAN2 |
|-----------|----------------|------|---------------------|
| Female    | 2.72473148     | 0.0449 |
| Male      | 3.19522401     |       |

| GENDER_S | TEACHER LSMEAN | Pr > |T| H0: LSMEAN1=LSMEAN2 |
|-----------|----------------|------|---------------------|
| Female    | 2.99151164     | 0.7841 |
| Male      | 2.92944386     |       |
Appendix J

Results of 1 Factor ANOVAs
Data Analysis Output from SAS Program

**One Factor ANOVAs:**

**Professor Gender-role Orientation**

### General Linear Models Procedure: Class Level Information

1: The SAS System

General Linear Models Procedure
Class Level Information

<table>
<thead>
<tr>
<th>Class Levels Values</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Androgynous</td>
<td>4</td>
</tr>
<tr>
<td>Feminine</td>
<td>4</td>
</tr>
<tr>
<td>Masculine</td>
<td>4</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>4</td>
</tr>
</tbody>
</table>

Number of observations in data set = 249

### Test of Model Significance: Progress Rating as Dependent Variable

2: The SAS System

General Linear Models Procedure
Dependent Variable: PROGRESS

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>27.88357530</td>
<td>18.23</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>245</td>
<td>124.87979457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>248</td>
<td>152.76336987</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square C.V. PROGRESS mean

<table>
<thead>
<tr>
<th>R-Square</th>
<th>C.V.</th>
<th>PROGRESS Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.182528</td>
<td>26.18608</td>
<td>2.72641863</td>
</tr>
</tbody>
</table>

### ANOVA Summary Table: Progress Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRO_BYS</td>
<td>3</td>
<td>27.88357530</td>
<td>18.23</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
### Marginal Means: Progress as Dependent Variable

#### The SAS System

General Linear Models Procedure
Least Squares Means

| GRO_BYS     | PROGRESS | Pr > |T|: H0: LSMEAN(i)=LSMEAN(j) | LSMEAN i/j | 1    | 2    | 3    | 4    |
|-------------|----------|------|--------------------------|------------|------|------|------|------|
| Androgynous | 3.15566969 | 1    | .                        | 0.0051     | 0.0005| 0.0001|      |      |
| Feminine    | 2.79430583 | 2    | 0.0051                   | .          | 0.5069| 0.0001|      |      |
| Masculine   | 2.70870365 | 3    | 0.0005                   | 0.5069     | .    | 0.0002|      |      |
| Undifferentiated | 2.22054148 | 4    | 0.0001                   | 0.0001     | 0.0002|       |      |      |

**NOTE:** To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
Test of Model Significance: Course Rating as Dependent Variable

IThe SAS System

General Linear Models Procedure

Dependent Variable: COURSE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>52.52097120</td>
<td>15.92</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>245</td>
<td>269.47902880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>248</td>
<td>322.00000000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square          | C.V.           | COURSE Mean
0.163109          | 34.95893       | 3.00000000

ANOVA Summary Table: Course Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRO_BYS</td>
<td>3</td>
<td>52.52097120</td>
<td>15.92</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Marginal Means: Course Rating as Dependent Variable

IThe SAS System

General Linear Models Procedure

Least Squares Means

| GRO_BYS | COURSE | Pr > |T| H0: LMEAN(i)=LMEAN(j) |
|---------|--------|------|------------------------|
| LMEAN   | i/j    | 1    | 2                       |
| Androgynous | 1.63076923 | 1     | 0.0086 0.0001 0.0001 |
| Feminine  | 1.13333333 | 2     | 0.0086 0.1049 0.0001 |
| Masculine | 2.82539683 | 3     | 0.0001 0.1049 0.0181 |
| Undifferentiated| 2.37704918 | 4     | 0.0001 0.0001 0.0181 |

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
Test of Model Significance: Instructor Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Dependent Variable: TEACHER

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>71.98276110</td>
<td>20.11</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>245</td>
<td>292.3246781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>248</td>
<td>364.30722892</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.197588
C.V. of TEACHER Mean: 36.92973

ANOVA Summary Table: Instructor Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRO_BTS</td>
<td>3</td>
<td>71.98276110</td>
<td>20.11</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Marginal Means: Instructor Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Least Squares Means

<table>
<thead>
<tr>
<th>GRO_BTS</th>
<th>TEACHER</th>
<th>Pr &gt;</th>
<th>H0: LSMEAN(i)=LSMEAN(j)</th>
<th>LSMEAN i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Androgynous</td>
<td>3.65384615</td>
<td>1</td>
<td>0.0442 0.0001 0.0001</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feminine</td>
<td>3.25832333</td>
<td>2</td>
<td>0.0442 0.0009 0.0001</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculine</td>
<td>2.59223810</td>
<td>3</td>
<td>0.0001 0.0009 0.1274</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>2.29508197</td>
<td>4</td>
<td>0.0001 0.0001 0.1274</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
One Factor ANOVAs:

Sex of Student

General Linear Models Procedure: Class Level Information

Class Level Information

Class Levels Values
GENDER_S 2 Female Male

Number of observations in data set = 126

Test of Model Significance: Progress Rating as Dependent Variable

ANOVA Summary Table: Progress Rating as Dependent Variable
**Marginal Means: Progress Rating as Dependent Variable**

The SAS System

General Linear Models Procedure

Least Squares Means

| GENDER | PROGRESS | Pr > |T| | HO: LSMEAN | LSMEAN1=LSMEAN2 |
|--------|----------|------|---|----------------|------------------|
| Female | 2.73499203 | 0.2774 |
| Male   | 2.58028388 |       |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
## Test of Model Significance: Course Rating as Dependent Variable

The SAS System 5

General Linear Models Procedure

Dependent Variable: COURSE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.43225358</td>
<td>0.31</td>
<td>0.5807</td>
</tr>
<tr>
<td>Error</td>
<td>124</td>
<td>174.77409562</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>125</td>
<td>175.20634921</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.002467, C.V.: 40.64905, COURSE Mean: 2.92063492

## ANOVA Summary Table: Course Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_S</td>
<td>1</td>
<td>0.43225358</td>
<td>0.31</td>
<td>0.5807</td>
</tr>
</tbody>
</table>

## Marginal Means: Course Rating as Dependent Variable

The SAS System 6

General Linear Models Procedure

Least Squares Means

| GENDER_S | COURSE | Pr > |T| H0: |
|----------|--------|------|-----|
|          | LSMEAN | LSMEAN1=LSMEAN2 |
| Female   | 2.98305085 | 0.5807 |
| Male     | 2.86567164 |      |
Test of Model Significance: Instructor Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Dependent Variable: TEACHER

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.00106208</td>
<td>0.00</td>
<td>0.9799</td>
</tr>
<tr>
<td>Error</td>
<td>124</td>
<td>207.20528712</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>125</td>
<td>207.20634921</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square  C.V.  TEACHER Mean
0.000005   44.26009  2.92063492

ANOVA Summary Table: Instructor Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_S</td>
<td>1</td>
<td>0.00106208</td>
<td>0.00</td>
<td>0.9799</td>
</tr>
</tbody>
</table>

Marginal Means: Instructor Rating as Dependent Variable

The SAS System

General Linear Models Procedure

Least Squares Means

<table>
<thead>
<tr>
<th>GENDER_S</th>
<th>TEACHER</th>
<th>Pr &gt;</th>
<th>\textit{T}</th>
<th>BO:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LSMEAN</td>
<td>LSMEAN1-LSMEAN2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2.92372881</td>
<td>0.9799</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.91791045</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
One Factor ANOVAs:

Sex of Professor

General Linear Models Procedure: Class Level Information

1The SAS System

General Linear Models Procedure
Class Level Information

Class Levels Values
GENDER_P 2 Female Male

Number of observations in data set = 249

Test of Model Significance: Progress Rating as Dependent Variable

1The SAS System

General Linear Models Procedure

Dependent Variable: PROGRESS

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>P Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.37391454</td>
<td>0.61</td>
<td>0.4370</td>
</tr>
<tr>
<td>Error</td>
<td>247</td>
<td>152.38945533</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>248</td>
<td>152.76336987</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square         C.V.    PROGRESS Mean
0.002448         28.80955 2.72641863

ANOVA Summary Table: Progress Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>P Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>0.37391454</td>
<td>0.61</td>
<td>0.4370</td>
</tr>
</tbody>
</table>
Marginal Means: Progress Rating as Dependent Variable

The SAS System

General Linear Models Procedure
Least Squares Means

GENDER | PROGRESS | Pr > |T| H0: LSMEAN | LSMEAN1=LSMEAN2

Female | 2.68843780 | 0.4370
Male   | 2.76595607 |
Test of Model Significance: Course Rating as Dependent Variable

The SAS System 5
General Linear Models Procedure
Dependent Variable: COURSE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>11.71556732</td>
<td>9.33</td>
<td>0.0025</td>
</tr>
<tr>
<td>Error</td>
<td>247</td>
<td>310.28443268</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>248</td>
<td>322.00000000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.036384
C.V.: 37.36029
COURSE Mean: 3.00000000

ANOVA Summary Table: Course Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>11.71556732</td>
<td>9.33</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

Marginal Means: Course Rating as Dependent Variable

The SAS System 6
General Linear Models Procedure
Least Squares Means

| GENDER_P | COURSE | Pr > |T| |80|         |
|----------|--------|------|---|---|----------------|
|          | LSMEAN | LSMEAN1=LSMEAN2 |
| Female   | 2.78740157 | 0.0025 |
| Male     | 3.22131148 |        |
### Test of Model Significance: Instructor Rating as Dependent Variable

The SAS System

General Linear Models Procedure

**Dependent Variable: TEACHER**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>19.2885083</td>
<td>13.81</td>
<td>0.0003</td>
</tr>
<tr>
<td>Error</td>
<td>247</td>
<td>345.018808</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>248</td>
<td>364.30722892</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square C.V. TEACHER Mean

- 0.052947 39.95759 2.95783133

### ANOVA Summary Table: Instructor Rating as Dependent Variable

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER_P</td>
<td>1</td>
<td>19.2885083</td>
<td>13.81</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

### Marginal Means: Instructor Rating as Dependent Variable

**1 The SAS System**

General Linear Models Procedure

Least Squares Means

<table>
<thead>
<tr>
<th>GENDER</th>
<th>TEACHER</th>
<th>Pr &gt;</th>
<th>T</th>
<th>H0:</th>
<th>LSMEAN</th>
<th>LSMEAN1=LSMEAN2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>2.68503937</td>
<td>0.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3.24180328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


192


Powell, G.N., & Butterfield, D.A. (1984). If "good managers" are masculine, what are "bad managers"? Sex Roles. 10, 477-484.


