Louisiana Tech University
Louisiana Tech Digital Commons

Doctoral Dissertations | Graduate School

Spring 5-2022

Assessing Gender Bias in Student Evaluations of Teaching in Collegiate Aviation

Kary M. Randall

Follow this and additional works at: https://digitalcommons.latech.edu/dissertations

ASSESSING GENDER BIAS IN STUDENT EVALUATIONS OF TEACHING IN COLLEGIATE AVIATION

by

Kary Munn Randall, A.S., B.S., M.A., M.A.

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education: Higher Education Administration

COLLEGE OF EDUCATION
LOUISIANA TECH UNIVERSITY

May 2022

LOUISIANA TECH UNIVERSITY

GRADUATE SCHOOL

March 10, 2022
Date of dissertation defense

We hereby recommend that the dissertation prepared by

Kary M. Randall, A.S., B.S., M.A., M.A.

entitled **Assessing Gender Bias in Student Evaluations of Teaching in Collegiate Aviation**

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Education, Higher Education Administration Concentration

Bryan McCoy
Supervisor of Dissertation Research

Dustin Hebert
Head of Curriculum, Instruction, and Leadership

Doctoral Committee Members:
Lori Jacques
Rick Shrubb

Approved:

Don Schillinger
Dean of Education

Approved:

Ramu Ramachandran
Dean of the Graduate School

GS Form 13a
(01/20)

ABSTRACT

Student evaluations of teaching (SET) are commonly used to assess teaching effectiveness and influence personnel decisions in higher education. This quantitative study sought to determine if gender and years of teaching experience were related to SET ratings for collegiate aviation faculty. Constructs evaluated related to gender stereotypes and consisted of expressiveness and immediacy as stereotypically female and professionalism and openness as stereotypically male. The overall rating was also analyzed as a fifth construct. Evaluation ratings from 54 participants associated with nine Aviation Accreditation Board International affiliated institutions were analyzed for the 2017 to 2020 academic years. Findings from the two-way MANOVA suggested no significant difference between ratings of the aviation faculty regardless of gender or years of teaching experience, Wilks' Λ=.860, $F(4, 47)$=1.908, p=.125, multivariate $\eta 2$=.140. A follow-up ANOVA of the between-subjects effect indicates no significant difference in ratings for expressiveness, immediacy, professionalism, openness, and overall based on gender and years of teaching experience. The lack of significant differences suggests that students in aviation do not associate these traits with the gender of aviation faculty. The similarities of aviation faculty in experience and personality type might be such that any gender differences are not evident in SET teaching ratings.

APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Dissertation. It was understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Dissertation. Further, any portions of the Dissertation used in books, papers, and other works must be appropriately referenced to this Dissertation.

Finally, the author of this Dissertation reserves the right to publish freely, in the literature, at any time, any or all portions of this Dissertation.

Author Kary Randall

Date 4/11/22

GS Form 14
(5/03)

DEDICATION

This dissertation is dedicated to the loving memory of my sister, Amy, who was enrolled in the Doctor of Education program two years before her passing. I did this for us.

TABLE OF CONTENTS

ABSTRACT iii

APPROVAL FOR SCHOLARLY DISSEMINATION iv

DEDICATION v

LIST OF TABLES ix

LIST OF FIGURES x

ACKNOWLEDGEMENTS xi

CHAPTER 1 INTRODUCTION 1

Background of the Problem 2

Statement of the Problem 2

Purpose of the Study 2

Significance of the Study 3

Methodology 3

Research Questions and Hypotheses 4

Definition of Key Concepts 5

CHAPTER 2 REVIEW OF LITERATURE 6

Theoretical Framework 7

Applications of Social Role Theory in Education 9

Validity of Student Evaluations of Teaching 11

Grade Bias 11

Course Workload/Difficulty Bias 13
Faculty Age Bias 14
Gender Bias in Student Evaluations of Teaching 17
Gender Affinity Effect 19
Gender Stereotyping 21
Gender Bias in SETs in Male-Dominated Academic Disciplines 25
Gender Bias in SETs in Collegiate Aviation 27
Summary 28
CHAPTER 3 METHODOLOGY 30
Research Questions 30
Population 31
Study Sample 31
Study Solicitation 33
Research Design 34
Data and Analysis 35
Data Collection 35
Data Analysis 39
Instrumentation 40
Assessing Content Validity 41
Threats to Validity 42
Inadequate Explication of Constructs 42
Low Statistical Power 42
Self-Reporting 42

Selection....43
CHAPTER 4 RESULTS....44
Descriptive Statistics....47
Data Analysis Procedures....50
Assumptions....52
MANOVA Results....56
Research Question 1 - Expressiveness....60
Research Question 2 - Immediacy....61
Research Question 3 - Professionalism....62
Research Question 4 - Openness....63
ANOVA Results....65
Research Question 5 –Overall Ratings....65
Summary....68
CHAPTER 5 DISCUSSION....69
Limitations....71
Delimitations....72
Implications for Practice....72
Implications for Future Research....72
Conclusion....75
REFERENCES....77
APPENDIX A AVIATION FACULTY SURVEY....87
APPENDIX B HUMAN USE EXEMPTION....90
APPENDIX C CHANGING EVALUATION TERMINOLOGY TO CONSTRUCTS....92

LIST OF TABLES

Table 1 *Distribution of Faculty by Participant Gender and Institution* 33

Table 2 *Descriptive Statistics: Distribution of Participant Groups*.....47

Table 3 *Descriptive Statistics Full Sample* 48

Table 4 *Descriptive Statistics by Gender* 48

Table 5 *Descriptive Statistics by Years of Experience* 49

Table 6 *Descriptive Statistics by Gender and Years of Experience* 50

Table 7 *Shapiro-Wilkes Test of Normality* 54

Table 8 *Correlations of Dependent Variables* 55

Table 9 *Multivariate Tests for Group* 57

Table 10 *Tests Between Dependent Variables Based on Gender* 58

Table 11 *Tests Between Dependent Variables Based on Years of Teaching Experience*.....58

Table 12 *Tests Between Dependent Variables Based on Gender and Years of Teaching Experience*.....59

Table 13 *Multiple Comparisons by Group and Dependent Variable Professional* 60

Table 14 *Shapiro-Wilk Test of Normality for Overall Rating*.....66

Table 15 *Two-way ANOVA Summary* 68

Table 16 *Descriptive Statistics by Construct and Participant Group*.....73

LIST OF FIGURES

Figure 1 *Outliers Based on Gender for Expressiveness* 37

Figure 2 *Outliers Based on Gender for Immediacy* 37

Figure 3 *Outliers Based on Gender for Professionalism* 38

Figure 4 *Outliers Based on Gender for Openness* 38

Figure 5 *Outliers Based on Gender Overall* 39

Figure 6 *Interaction Between Gender and Years of Teaching Experience for Expressiveness* 61

Figure 7 *Interaction Between Gender and Years of Teaching Experience for Immediacy* 62

Figure 8 *Interaction Between Gender and Years of Teaching Experience for Professionalism* 63

Figure 9 *Interaction Between Gender and Years of Teaching Experience for Openness* 64

Figure 10 *Visual Representation of Interactions Between Factors* 67

ACKNOWLEDGMENTS

First, I would like to acknowledge that this would not have been possible without the strength, focus, and endurance given to me by God. I am forever grateful for the experience and the personal growth I have gained on this journey. I would also like to thank Louisiana Tech University for this opportunity and thank my committee members, Dr. Bryan McCoy, Dr. Lorraine Jacques, and Dr. Richard Shrubb. Without your support and faith in me, completing this chapter would have been more arduous.

A special thanks to Dr. Jacques for her unwavering guidance and constant encouragement throughout this process. I cannot say enough about her kindness, loyalty, and friendship. She taught me the fine points of research and gave me the confidence to use them. She was an excellent mentor who was firm when it was necessary and compassionate when I needed it most. I am a better researcher and instructor because of her. I would also like to thank my chair, Dr. McCoy. I would never have made it to this point without his direction and great sense of humor. Also, thank you to Dr. Shrubb for being a calming voice during all the storms along the way.

I would also like to thank my friends and colleagues who tolerated listening to the constant talk of research methodology, even when they did not want to. I would like to recognize Jordan Lyons, who shared this experience with me and never tired of discussing my concerns and worries. I also cannot go without mentioning Kinbur Blackmon Thornburg, who has stood next to me through the last year of my journey.

I appreciated her daily calls to listen and offer her words of encouragement. I know she felt the successes of this experience as much as I did. I am most appreciative of my family. I thank my husband, Jim, and mother, Karen, for their love, support, and patience. They both stepped in to take care of things when I could not. I love you and am blessed to have you in my life.

CHAPTER 1

INTRODUCTION

Higher education institutions frequently use student evaluations of teaching (SET) to assess faculty teaching effectiveness (Algozzine et al., 2004; Galbraith et al., 2011; Wachtel, 1998). SETs can be traced to the late 1920s, when higher education institutions implemented them to evaluate faculty performance to improve teaching quality (Algozzine et al., 2004). Today, SETs are used as part of a cumulative means of evaluating faculty competence (Arreolla, 2000; Centra, 1979; Seldin, 1999) and are commonly used to make personnel decisions such as tenure and promotion (Cashin, 1999; Marcham et al., 2020; Seldin, 1999). The validity and reliability of SETs are criticized for the variables affecting their efficacy (Bharadwaj et al., 1993; Blackhart et al., 2006; Boring et al., 2016; Greenwald & Gillmore, 1997). Variables found to influence SET ratings include course grades (Blackhart et al., 2006; Boring et al., 2016; Greenwald & Gillmore, 1997), course workload/difficulty (Centra, 2003; Thornton et al., 2010), faculty age (Joye & Wilson, 2015; Stonebraker & Stone, 2015; Wilson et al., 2014), and faculty gender (Basow & Silberg, 1987; Chamberlin & Hickey, 2001; Kreitzer & Sweet-Cushman, 2021; MacNell et al., 2015). This dissertation examined gender bias in SETs for collegiate aviation faculty.

Background of the Problem

Students rate faculty differently based on gender, including the gender affinity effect (Bachen et al., 1999; Boring, 2016; Chamberlin & Hickey, 2001) and gender stereotyping (Basow, 2000; Bennett, 1982; Kierstead et al., 1988). These biases emphasize a limitation of SETs that could adversely affect female faculty (Kreitzer & Sweet-Cushman, 2021). On average, female faculty rated lower than male faculty on overall satisfaction (Basow & Silberg, 1987; Kreitzer & Sweet-Cushman, 2021).

Gender bias was found in male-dominated disciplines such as engineering, business, and natural sciences (Basow & Silberg, 1987; Centra & Gaubatz, 2000; Narayanan et al., 2014). Females were rated lower in engineering and business (Basow & Silberg, 1987; Narayanan et al., 2014) or higher in stereotypically female areas (Centra & Gaubatz, 2000). Although Marcham et al. (2020) did not find disparities in online aviation courses, other instructional formats might have different findings.

Statement of the Problem

Men make up 85% of the faculty in collegiate aviation programs and 80% of the overall aviation workforce (Lutte, 2021). Given the findings in other male-dominated academic disciplines, female faculty may be disadvantaged when students evaluate their teaching effectiveness based on gender (Kreitzer & Sweet-Cushman, 2021). More needs to be known about how students evaluate teaching in collegiate aviation.

Purpose of the Study

The purpose of this quantitative study was to determine if gender and years of teaching experience are related to student evaluations of teaching for collegiate aviation

faculty. These constructs are related to gender norms and stereotypes, including expressiveness, immediacy, professionalism, openness, and the overall SET rating.

Significance of the Study

This research study contributed to the existing body of knowledge on the validity of SETs used in colleges and universities for measuring teaching effectiveness. Additionally, this study adds to the research on gender in collegiate aviation programs, specifically how aviation students evaluate the teaching effectiveness of their professors. The findings of this study may aid similar collegiate programs.

Methodology

Quantitative methods helped determine how gender and years of teaching experience were related to SETs. Participants from nine Aviation Accreditation Board International (AABI) affiliated institutions provided SET ratings from the 2017-2018, 2018-2019, and 2019-2020 academic years. Additionally, participants completed a questionnaire that aided in collecting information on their gender, institutional ranking, years of experience instructing, tenure status, and primary area of instruction. The current researcher used a matched pairs design by grouping participants by gender (male, female) and years of teaching experience (0-6 years, 7+ years) to eliminate the extraneous variable of experience.

Version 27 of the Statistical Package for Social Sciences (SPSS) aided in analyzing the student ratings of expressiveness, immediacy, professionalism, openness, and the overall SET rating for each participant grouping. A factorial multivariate analysis

of variance (MANOVA) determined the differences between participant groups with multiple independent and dependent variables (Mertler & Vannatta, 2013).

Research Questions and Hypotheses

RQ1: What difference, if any, exists between student ratings of expressiveness for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student expressiveness ratings for female faculty and male faculty in collegiate aviation programs.

RQ2: What difference, if any, exists between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs.

RQ3: What difference, if any, exists between student ratings of professionalism for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of professionalism for male faculty and female faculty in collegiate aviation programs.

RQ4: What difference, if any, exists between student ratings of openness for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of openness for male faculty and female faculty in collegiate aviation programs.

RQ5: What difference, if any, exists between the overall student ratings of female faculty compared to that of male faculty in collegiate aviation programs?

H_0: There is no difference between the overall student ratings of female faculty and male faculty in collegiate aviation programs.

Definitions of Key Concepts

AABI: Aviation Accreditation Board International is an accreditation system that recognizes collegiate aviation programs that maintain high-performance standards in quality, performance, and integrity (AABI, 2021).

Aviation program: A group of aviation courses that result in a degree with a defined area of specialization (AABI, 2021).

Equity Bias: Instructor variables beyond their control that influence outcomes (Kreitzer & Sweet-Cushman, 2021).

Experienced Faculty: Any faculty member with 7+ years of teaching experience (Iglesias-Martinez et al., 2014).

Novice Faculty: Any faculty member with 0-6 years of teaching experience (Iglesias-Martinez et al., 2014).

SET: Student evaluation of teaching (Mitchell & Martin, 2018).

CHAPTER 2

REVIEW OF LITERATURE

Research on the influence of faculty gender on student evaluations of teaching (SETs) is the focus of this review of the literature. Research on faculty gender in SETs is inconclusive with mixed results on its effect on student ratings (Boring, 2016; Chamberlin & Hickey, 2001; Feldman, 1993; MacNell et al., 2015). Due to the use of SETs in making personnel decisions regarding faculty tenure and promotion (Arreolla, 2000; Cashin, 1999; Centra, 2003; Hornstein, 2017), gender bias is concerning when SET ratings negatively influence faculty based on a characteristic outside of their control (Kreitzer & Sweet-Cushman, 2021). SETs are a valid and reliable means of measuring teaching effectiveness and improving teaching performance (Cohen, 1980; Feldman, 1992; Marsh, 1987); however, variables such as final or expected course grades (Bharadwaj et al., 1993; Blackhart et al., 2006; Boring et al., 2016; Greenwald & Gillmore, 1997), course workload/difficulty (Centra, 2003; Thornton et al., 2010), faculty age (Joye & Wilson, 2015; Stonebraker & Stone, 2015; Wilson et al., 2014), andfaculty gender (Bachen et al., 1999; Basow & Silberg, 1987; Boring, 2016; Chamberlin & Hickey, 2001; Kierstead et al., 1988; MacNell et al., 2015) were found to influence SET ratings.

Educational databases used to identify relevant studies consisted of EBSCO, ERIC, JSTOR Journals, SocIndex, Communication & Mass Media Complete, APA

PsycInfo, Professional Development Collection, Directory of Open Access Journals, Psychology and Behavioral Science Collection, SCOPUS, MLA International Bibliography, Business Source Complete, Academic Search Ultimate, and Google Scholar. Key terms used consisted of (a) gender bias in student evaluations of teaching, (b) gender bias, (c) collegiate aviation, (d) gender stereotypes, (e) teacher evaluations, (f) faculty performance, (g) higher education, (h) student evaluations, (i) gender in colleges and universities, (j) gender role, (k) performance evaluation, (l) teaching effectiveness, and (m) male-dominated disciplines. Additionally, research articles from University Aviation Association, Embry-Riddle Aeronautical University, and relevant articles were used. Articles that were not accessible or did not address student evaluations in a higher education setting were rejected. Approximately 16 quantitative and qualitative studies related to the influence of instructor gender on SETs were included in this literature review.

Five sections were used to review the literature. They consisted of (a) review of social role theory as the theoretical framework for this research, (b) relevant literature on the biases in SETs of collegiate instructors, (c) literature on gender biases in SETs, (d) literature related to gender biases in SETs in male-dominated disciplines, and (e) a review of relevant literature on gender biases in SETs of collegiate aviation faculty.

Theoretical Framework

Social role theory was the theoretical framework for this study. Social role theory is an approach to understanding the behavioral differences of men and women (Eagly, 1987). Social role theorists believe the origin of adult gender behaviors is a product of society's perceptions of gender roles. Men and women behave according to what they

observe as their societal roles at work and home. In other words, men and women model the behaviors they observe in society.

Alice Eagly developed social role theory in the 1980s after discovering a lack of scholarship on the social-psychological view of gender differences and the adaptability of gender roles (Eagly & Wood, 2012). Social role theory evolved from role theory, which is a theory that helps explain how peoples' expectations guide behavioral norms appropriate for various societal roles (Barnett, 2014). Role theory provided a framework that social role theorists expanded to explain changes in gender roles and behaviors in society. (Eagly & Wood, 2012). Unlike role theorists, social role theorists believe that as societal roles of men and women evolve, so do their behaviors. Social role theory also incorporates ideas presented in other sociological theories such as social-learning theory and biological theories. Childhood socialization theorists believe children learn their gender roles from tutoring they receive from parents, teachers, and other society members (Little, 2016). Social role theorists believe adult behaviors are learned during childhood and throughout life; however, adult gender behaviors are a malleable product of peoples' social observations (Eagly, 1987).

Male and female physical attributes also represent one factor that has influenced the development of social roles at work and home (Eagly & Wood, 2012). For example, the belief that men are bigger and stronger than women make them better suited to fill the role of provider for the household (Eagly, 1987). On the other hand, the belief that women are petite in size and more nurturing makes them better suited to care for the family and household (Eagly & Wood, 2012). Today, however, domestic and occupational roles do not rely as much on the physical attributes of men and women. This

change creates more opportunities for women in the workforce, thus resulting in more women working outside the home (Eagly, 1987).

Applications of Social Role Theory in Education

Social role theory explains gender roles in society based on observed behaviors that become stereotypical of men and women (Eagly & Wood, 2012). When men and women do not fit the roles hey are expected to fill, society can experience backlash (Froehlich et al., 2022). Froehlich et al. (2022) conducted three studies to determine how science, technology, engineering, and mathematics (STEM) disciplines were stereotyped by German and Japanese university students and explored societal repercussions and psychological consequences on the emotions and motivation of female STEM students. Germany and Japan were appropriate geographical locations because they have top rankings in STEM, and in both cultures, people are sensitive to societal backlash.

In the first two studies, Froehlich et al. (2022) collected data from two different groups of Japanese and German university students to determine gender stereotypes surrounding math and academic abilities. In the first study, participants listed and rated the valence of stereotypical statements about women's and men's math and intellectual abilities. Of the over 1,140 statements, a mixed-method analysis revealed participants rated women more negatively than men on math abilities, which indicated math abilities were a stereotypically male competence.

In the second study, Froehlich et al. (2022) collected data from a large sample that categorized words pairing male/female with science/liberal arts to examine the explicit and implicit gender-science stereotypes. ANOVA results revealed men were associated more than women with science, while women were associated more than men with liberal

arts. The findings of both studies suggested a negative gender stereotype of women's STEM abilities, perhaps explaining why there are fewer females in STEM fields.

Froehlich et al. (2022) conducted a third study to examine expected backlash for female STEM students through scenario-based questions. German and Japanese physics, engineering, and computer science students responded to a two-part online questionnaire. The first part consisted of 10 items used to determine the self-construal nature of each participant. The second part asked participants to describe the reactions of a stranger when discussing their field of study. Froehlich et al. (2022) hypothesized that women would expect adverse reactions, consequently affecting their emotions and motivation.

Along with writing down the anticipated responses, participants rated the valence of the reactions (Froehlich et al., 2022). Froehlich et al. (2022) found female participants expected strangers to be surprised by their chosen fields of study but not in a negative way. Female participants described surprised responses from strangers, which were more pronounced than those described by male participants. However, female STEM students also expected the strangers to rate them lower in traits such as gentleness, affection, and sympathy. The lower ratings in these stereotypically feminine traits indicated a subtle expected backlash. Though subtle, the backlash could negatively affect female STEM students when they decide not to pursue a STEM career after graduating from college (Froehlich et al., 2022).

Social role theory might explain why there are so few females in the male-dominated field of aviation (Ison, 2010; Luedtke, 1993; Lutte, 2021). Social role theory provides a framework for understanding how faculty gender influences SET ratings for collegiate aviation faculty.

Validity of Student Evaluations of Teaching

The validity of SET ratings becomes questionable when they measure variables unrelated to some aspect of teaching effectiveness (Cashin, 1999; Marsh, 1987). Variables found to influence SET ratings of collegiate instructors are final or expected course grade (Bharadwaj et al., 1993; Blackhart et al., 2006; Boring et al., 2016; Greenwald & Gillmore, 1997), course workload/difficulty (Centra, 2003; Thornton et al., 2010), faculty age (Joye & Wilson, 2015; Stonebraker & Stone, 2015; Wilson et al., 2014), and instructor gender (Bachen et al., 1999; Basow & Silberg, 1987; Boring, 2016; Chamberlin & Hickey, 2001; Kierstead et al., 1988; MacNell et al., 2015). Below is a literature review on SET biases, which emphasizes those biases related to instructor gender.

Grade Bias

Grade bias occurs when SET ratings are influenced by the awarded grade or expected grade for a course and not instructor effectiveness (Boring et al., 2016; Griffin et al., 2014). While Marsh and Roche (2000) found the relationship between expected course grades and SETs was stable over time, faculty and administrators believe that inflated grades lead to higher SETs. The view that collegiate faculty can buy higher SET ratings by awarding higher grades is not without merit (Centra, 2003).

Blackhart et al. (2006) found grade bias a predictor of SET ratings of instructors in the psychology department at Florida State University. Blackhart et al. (2006) analyzed over 9,000 SETs to determine what variables influenced student ratings. Blackhart et al. (2006) conducted multiple regression analyses between variables and discovered a significant correlation between SET ratings and average grades. The higher the grades

awarded by instructors, the higher student evaluations of faculty. The conclusion that students rated course instructors higher because of their higher grades was unclear. It is possible that student grades reflected the instructor's teaching effectiveness and the level of engagement students had in the course.

Adding clarity to the relationship between course grades and SETs, Boring et al. (2016) analyzed the influence of expected course grades on SET ratings at a French university. Boring et al. (2016) collected data from over 23,000 SETs and paired average SET scores with average interim grades. Boring et al. (2016) theorized that the average interim grade served as an indicator of future grades, establishing an expected grade. Boring et al. (2016) found a positive correlation between expected grades and SET ratings for most disciplines evaluated in her study. This finding indicated that SET ratings measured student satisfaction and grade expectations but not instructor teaching effectiveness.

Bharadwaj et al. (1993) investigated how student perceptions of collegiate instructors changed due to grades during a semester. Bharadwaj et al. (1993) collected data throughout the semester from 73 undergraduate students enrolled in a marketing class at a large university. Each student completed a survey rating the instructor on (a) overall satisfaction, (b) the role of the instructor, (c) teacher-student relationship, and (d) course quality. The students completed the surveys on four separate occasions (a) three weeks into the semester, (b) a week after the first exam, (c) before the final grade was issued, and (d) after the final grade was issued. Bharadwaj et al. (1993) performed a MANOVA to analyze student ratings for four teaching constructs. The results indicated student ratings changed over the semester with scores peaking before the

final grade was issued and dropping for all four constructs once the final course grades were given. This result indicated the final course grades negatively affected SET ratings. Bharadwaj et al. (1993) suggested instructors would receive higher evaluation ratings if they administered teacher performance evaluations before posting final grades.

Greenwald and Gillmore (1997) found the instructor grading policy influenced the SET ratings of faculty at the University of Washington. With individual courses used as the unit of analysis, Greenwald and Gillmore (1997) collected evaluation rating data from just under 900 courses university-wide. The evaluation form included seven items on student learning outcomes: one item on instructor appreciation, one item on the student's expected course grade, and one item that requested the number of hours per week the student spent doing course activities. Greenwald and Gillmore (1997) used the Grading Leniency Model to determine the relationship between evaluation ratings, expected grades, and course workload. By measuring course workload and finding courses with lighter workloads also had higher grades, Greenwald and Gillmore (1997) demonstrated the influence of grading leniency on student ratings.

Course Workload/Difficulty Bias

Centra (2003) found that course difficulty/workload was moderately related to SETs. Centra (2003) analyzed SET data from 55,000 courses for the 1995-1999 academic years. Among other variables, students rated expected grades and course workload/difficulty. Centra (2003) calculated course workload/difficulty by averaging responses from SET items on a) course preparation, b) workload compared to other courses, and c) the pace the instructor covered material. Using multiple regression analysis, Centra (2003) found expected grades did not affect student evaluation ratings;

however, the findings for workload/difficulty were more interesting. Centra (2003) discovered courses taught at an appropriate level scored higher than courses thought too easy or too difficult. These findings suggested course workload influenced SET ratings.

Thornton et al. (2010) found mixed results for professors teaching challenging courses from a small business school at a southeastern university. While investigating factors influencing SET ratings on the Student Instructional Report (SIR) II instrument, Thornton et al. (2010) collected evaluation reports from 80 students. Thornton et al. (2010) used a multiple regression model to determine the relationships between the mean overall evaluation and 13 independent variables from the SIR II, including grading, workload, pace, and effort. The results indicated that grading, workload, and pace affected the overall evaluation score with student effort negatively related to the overall evaluation score. The results showed that instructors who were considered challenging required more student effort and had lower overall evaluation ratings than those who were teaching easier courses.

Faculty Age Bias

Wilson et al. (2014) examined the effects of age and professor gender on SETs of 231 undergraduate students from a southeastern university. Wilson et al. (2014) hypothesized students would rate younger professors higher than older professors. They further hypothesized that while younger female professors would rate higher than more senior female professors, the ratings between younger and older male professors would not change. Wilson et al. (2014) presented each participant with one of four randomly assigned black-and-white photos of a man or a woman as a younger or older adult. Participants then rated the image based on how they believed the person would behave as

a professor. Afterward, participants completed a survey designed to measure (a) the professor's encouragement of questions, (b) expectations of good work, (c) workload, (d) organization, (e) explanation of concepts, (f) friendliness, and (g) being a good teacher. Students also completed a 34-question Professor-Student Rapport Scale to assess their perceived rapport with the professor. Participants then indicated how old they assumed the professor to be and their level of attractiveness.

Wilson et al. (2014) conducted a MANOVA to analyze professor gender and age on the survey ratings, the rapport scale ratings, and attractiveness. Results revealed that age influenced the student ratings for male and female professors. Students rated the younger professor higher for (a) teacher encouragement, (b) expectations of good work, (c) workload, (d) friendliness, (e) student rapport, and (f) attractiveness. Students rated the young female professor as more organized and attractive than the young male and older female professors. As expected, the more senior female professor was rated more negatively than the younger female professor was; however, age was not related to the ratings of male professors.

Joye and Wilson (2015) analyzed student evaluations to determine if gender and age affected the SET ratings of 340 student participants from psychology courses at a southeastern university. Joye and Wilson (2015) hypothesized that the younger female professors would rate higher on attractiveness and rapport. Like Wilson et al. (2014), Joye and Wilson (2015) presented participants with a black-and-white photo of a man or a woman as a younger or older adult. Participants then listened to a 3-minute history lecture introduced by a gender-ambiguous voice, followed by a quiz. Participants completed a seven-item assessment survey and Professor-Student Rapport Scale to rate

how they believed the pictured individual would behave as a professor for (a) encouraging questions, (b) having expectations of good work, (c) maintaining an adequate workload, (d) keeping lessons organized, (e) explaining concepts, (f) upholding friendliness, and (g) teaching well. Additionally, participants rated how they believed their rapport with the pictured professor would be, how attractive the pictured professor was, and the age they thought the pictured individual to be.

A MANOVA comparing gender and age of professor on the seven-item assessment, rapport scale, and attractiveness revealed the perceived age of the instructor affected the students' perceptions of professor-student rapport, attractiveness, and quiz grades (Joye & Wilson, 2015). Joye and Wilson (2015) found students rated younger professors as more attractive and assumed they would have a better rapport; however, they scored higher on the quiz if they believed the lecture came from an older professor.

Stonebraker and Stone (2015) analyzed the Rate My Professor (RMP) ratings of 3,600 tenure-track professors to determine if age affected how students perceived their teaching effectiveness. Stonebraker and Stone (2015) collected age and tenure information from 58 institutions' websites and the RMP ratings for helpfulness, clarity, and ease. Using a regression model then clustering the standard errors at the institutional level, Stonebraker and Stone (2015) found that age negatively affected teacher quality ratings for instructors over 45 regardless of gender, academic discipline, and type of institution. The findings further revealed that factors such as attractiveness and ease of the instructor improved their overall effectiveness ratings. For instructors rated as attractive, the effects of age on RMP ratings were irrelevant. The findings indicated that the instructor's age influenced students' perceptions of teaching quality.

Gender Bias in Student Evaluations of Teaching

The validity of SETs becomes questionable when personal traits such as faculty age and gender are evaluated instead of teaching effectiveness (Arbuckle & Williams, 2003; Kreitzer & Sweet-Cushman, 2021). While studies found no differences in SET ratings among male and female faculty (Bennett, 1982; Feldman, 1992; Marcham et al., 2020), a vast number of studies found faculty gender influenced SET ratings (Bachen et al., 1999; Basow & Silberg, 1987; Boring, 2016; Chamberlin & Hickey, 2001; Kierstead et al., 1988; MacNell et al., 2015; Mitchell & Martin, 2018).

Demonstrating that gender and age were related to SET ratings, Arbuckle and Williams (2003) investigated student perceptions of expressiveness, faculty gender, and age. Students from six introductory psychology classes watched a 35-minute audiovisual slide lecture with a gender-neutral stick figure and voice. Students then completed a SET questionnaire with one of four gender and age descriptions of the presenter. The professor categories consisted of a female under 35, a male under 35, a female over 55, and a male over 55. Arbuckle and Williams (2003) examined gender and age categories to determine if they would prompt students to use gender and age stereotypes in answering the SET questions. Arbuckle and Williams (2003) found students rated the young male professor higher than any other category, suggesting students expected the college professor to be young and male.

MacNell et al. (2015) investigated how students enrolled in online social sciences courses rated the teaching effectiveness of their collegiate instructors based on what they perceived the instructors' genders to be. MacNell et al. (2015) hypothesized that there would be no difference in the ratings of instructors regardless of gender. Forty-three

students were randomly assigned to online discussion groups. A male and female instructor were assigned to instruct one class as their actual genders and a second class as the opposite genders. The instructors had no contact with the students outside the online environment. Each instructor taught the course similarly by presenting similar credentials, covering the same assignments, and returning work at the same rate. Participants submitted instructor ratings for six teaching effectiveness traits, six interpersonal traits, and overall instructor quality. Using a 2 X 2 experimental design, MacNell et al. (2015) made SET rating comparisons across the instructors' actual and perceived genders. The findings indicated a significant difference in how students rated the perceived male and female instructors. The perceived male instructor rated higher than the female instructor in (a) professionalism, (b) promptness, (c) fairness, (d) respectfulness, (e) enthusiasm, (f) praise, and (g) overall instructor quality. The results of this study supported the existence of gender bias in SETs.

Boring et al. (2016) used the same dataset collected by MacNell et al. (2015) to determine if a non-parametric test would yield the same results. Boring et al. (2016) used permutation tests to determine if perceived instructor gender influenced SET ratings. The results revealed an association between instructor gender and SET ratings. Boring et al. (2016) found males rated higher overall and in the areas of (a) fairness, (b) promptness, (c) giving praise, (d) enthusiasm, (e) respect, and (f) caring. Although the significance of the non-parametric tests was smaller than those reported by MacNell et al. (2015), the results suggested male faculty rated higher than female faculty.

Mitchell and Martin (2018) modeled their study after the MacNell et al. (2015) study by comparing the SET ratings for an identical online course taught by one male and

one female instructor. Mitchell and Martin (2018) used introductory political sciences courses with identical lectures, assignments, and content. Upon completing the online courses, students answered a 23-question evaluation consisting of categories for (a) instructor/course, (b) course, (c) technology, and (d) administrative categories. Mitchell and Martin (2018) hypothesized that the male instructor would rate higher than the female instructor for SET categories related to instructor/course, course, and technology. Because the administrative questions were not associated with the course content, they hypothesized that the male and female instructors would rate equally. The results revealed that the male instructor rated significantly higher than the female instructor for the instructor/course, course, and technology categories but not significantly different for the administrative category. The findings indicated that the female instructor rated lower because of the instructor's gender. Mitchell and Martin (2018) further concluded the observed similarities in administrative ratings indicated students carefully read each question before answering.

There were exceptions to these gender biases that fell into two categories of gender affinity effect (Bachen et al., 1999; Basow & Silberg, 1987; Chamberlin & Hickey, 2001) and gender stereotyping (Bachen et al., 1999; Basow, 2000; Bennett, 1982; Rubin, 1980). The following section will describe these in more detail.

Gender Affinity Effect

The gender affinity effect in SETs occurs when students rate their same-gendered instructor higher regardless of teaching effectiveness (Kreitzer & Sweet-Cushman, 2021). The gender affinity effect might explain the findings in a study by Bachen et al. (1999) to determine if student assessments of faculty were influenced by gender schema. Bachen et

al. (1999) asked 486 undergraduate students from a mid-sized private university to choose professors from their past or present and rate them on five teaching dimensions consisting of (a) caring-expressiveness, (b) professional-challenging, (c) interactive, and (d) organized. Bachen et al. (1999) performed a MANOVA finding significant effects between student gender and faculty gender for all teaching dimensions measured. Follow-up ANOVAs revealed female students rated female faculty significantly higher on all five teaching dimensions. The female students rated female faculty higher in the traditionally female traits of caring-expressive, interactive, and easy-going, and traditionally male features of professional-challenging and organized. Male students, however, rated male and female faculty similarly in all five areas. The findings indicated that student gender influenced SET ratings with the female students accounting for the strong interaction. Bachen et al. (1999) concluded that female students rated female faculty higher because they related more to female faculty than male faculty.

Chamberlin and Hickey (2001) investigated if male and female instructors were rated differently by male and female students. Chamberlin and Hickey (2001) hypothesized that female students would rate faculty based on gender stereotypes, and male students would rate faculty based on gender. They collected questionnaire data from 198 undergraduate students enrolled in introductory sociology and anthropology courses. Students rated instructors on 12 teaching attributes. Chamberlin and Hickey (2001) found a significant difference for nine teaching attributes, indicating that students rated male and female instructors differently. Chamberlin and Hickey (2001) cross-tabulated instructor gender by each teaching attribute, controlling for student gender. The results indicated female students rated female instructors significantly higher in most teaching

attributes, suggesting that female students were more likely to evaluate faculty differently than male students. Although the reason was unknown, female students preferred instructors like themselves.

Basow and Silberg (1987) found male students gave male professors higher ratings than female professors on SET questions related to (a) scholarship, (b) organization/clarity, (c) instructor-group interaction, (d) instructor-individual student interaction, (e) dynamism, and (f) overall teaching ability. Although female students rated male and female professors similarly, they rated female professors lower than male professors on individual student interaction, dynamism, and overall teaching ability. Basow and Silberg (1987) suggested that the lower scores for female professors for instructor and student interaction might have been because the students expected female professors to be accessible.

Gender Stereotyping

Gender stereotypes emerge when observed male and female behaviors become shared generalizations made by society (Eagly & Wood, 2012). These behavior generalizations become synonymous with perceived gender social roles and develop into male and female personality traits (Eagly & Wood, 2012; Verniers & Martinot, 2015). Where women are stereotypically caring, supportive, kind, and concerned for others, men are stereotypically more dominant, assertive, decisive, and independent (Eagly & Wood, 2012; Heilman, 2001). Stereotyping as a means of sorting information is beneficial for efficiency; however, stereotyping people can lead to biases that wrongfully affect people when decisions are made solely on one's gender (Heilman, 1997).

Basow (2000) examined student descriptions of their professors to determine which traits they most valued in a collegiate setting. Ninety-eight students from a small liberal arts college described their best and worst professors and rated them using the Bem Sex-Role Inventory. Basow (2000) used the Bem Sex-Role Inventory to evaluate traits considered either stereotypically male or stereotypically female and found students valued different qualities of their male and female professors. Students often described the "best" female professors as helpful and approachable and the "best" male professors as organized and open-minded. The disparate qualities students valued in male and female professors suggested gender was a factor in student perceptions of faculty performance. Basow (2000) concluded that more attention was needed on gender and teaching evaluations.

Basow's (2000) findings were like those found by Bachen et al. (1999) when nearly 500 undergraduate students were asked to describe the differences between their male and female professors. Students emphasized female traits more often, noting their approachability, supportiveness, and enthusiasm. However, students criticized female professors when they lacked these nurturing traits. Basow (2000) concluded that students held different expectations of the teaching styles of male and female professors.

Rubin (1980) analyzed student responses to determine which teaching traits were ideal for male and female professors. Rubin (1980) separated 127 undergraduate students from a small mid-western university into three groups. Using a list of 34 teaching traits, Rubin (1980) asked one group of students to identify traits ideal for a male professor, a second group to identify traits ideal for a female professor, and a third group to determine the ideal traits in general. Separated into five manageable categories, Rubin (1980)

compared the percentage of student responses. The results indicated students expected female professors more than male or unspecified professors to possess qualities associated with the nurturing trait. Students further expected male professors, more than female or unspecified professors, to have characteristics associated with open-mindedness. Rubin's (1980) results suggested students hold disparate expectations for male and female professors that conform to perceived sex roles.

Bennett (1982) also concluded students expected specific stereotypical behaviors from their female professors. Bennett (1982) examined student responses to determine if male and female instructors were rated similarly on formal evaluations and if students received more personal contact with their female instructors. Bennett (1982) asked 253 undergraduate students at a liberal arts college to complete a formal teaching performance evaluation and a student-instructor contact indicator questionnaire. Students rated their instructors on their (a) nonauthoritarian interpersonal style, (b) charisma, (c) self-assurance, and (d) instructional approach. Students further reported (a) the level of contact they had with their instructors, (b) the context of that contact, and (c) how freely they felt contacting their instructors. Bennett (1982) calculated bivariate correlations for male and female instructors for each performance item revealing no evidence of significant differences in ratings on the formal evaluation. However, students reported receiving more personal contact with female instructors than male instructors. Regardless of gender, students reported more visits to their female instructors' offices, being more comfortable contacting their female instructors at home, and having more personal discussions with their female instructors; however, they failed to rate female

faculty higher than male faculty on the formal evaluation items related to availability. Bennett (1982) concluded that female faculty are expected to be accessible.

Kierstead et al. (1988) examined how warmth and friendliness affected student ratings of male and female faculty. They hypothesized that instructors, particularly female instructors, who smiled and interacted with students more would receive higher student ratings of teaching effectiveness. Kierstead et al. (1988) separated 40 students into two groups who then read a description of a course professor's behavior, availability, and a summary of the amount of out-of-class contact the professor had with students. Half the students were told the professor was male, and the other half were told the professor was female. After reading the description, students evaluated the professor. An ANOVA revealed the male instructor received higher ratings than the female instructor regardless of the amount of social interaction he/she provided the students. However, the female instructor was rated higher when she increased the social interaction with students and was rated lower when she decreased the social interaction with students. These findings indicated students rewarded the female instructor when she met their expected gender stereotype and punished her when she did not.

In a second experiment, Kierstead et al. (1988) examined the effects of smiling on student ratings of instructors. Participants watched a presentation of a pre-recorded lecture. Half the students received the female's class and half the male's class. Throughout their presentations, groups of female and male lecturers smiled or did not smile. Upon completion of the presentation, students completed an evaluation of teaching performance. An ANOVA revealed that male faculty rated higher when not smiling at students; however, the smiling female faculty received higher ratings than the unsmiling

female faculty. The findings of both studies indicated faculty who fit into their social role expectation were rewarded with higher evaluation scores.

Boring (2016) examined SET ratings to determine how gender bias influenced teaching effectiveness. Boring (2016) collected 20,197 SET ratings from first-year students over 5 academic years. Students rated faculty on four dimensions (a) course content, (b) assignments, (c) delivery style, and (d) professor's knowledge. Using a regression model to determine the relationship between student and faculty genders and SET ratings for each teaching attribute, Boring (2016) found male and female students gave higher ratings to male professors on teaching attributes associated with male stereotypy for leadership and knowledge. Students rated female professors higher on clarity and usefulness of feedback on assignments, which may be related to the female stereotype of warmth and nurturing. Overall, the findings of this study suggested that stereotypes held by students may influence the ratings given on SETs and may also indicate students have different standards for male and female professors.

Gender Bias in SETs in Male-Dominated Academic Disciplines

Centra and Gaubatz (2000) analyzed data from 741 classes from 21 institutions to determine the existence of gender bias in student evaluations across eight academic disciplines, including traditionally male-dominated fields of business, natural sciences, and technology. Centra and Gaubatz (2000) collected student ratings from within classes and across classes to make comparisons of student ratings within each grouping. Centra and Gaubatz (2000) analyzed student ratings for eight teaching attributes consisting of (a) course organization and planning, (b) communication, (c) faculty/student interaction, (d) assignments, exams, and grading, (e) course outcomes, (f) student effort and

involvement, (g) course difficulty and workload, and (h) overall. A MANOVA comparing differences in mean ratings across classes revealed female and male students favored the same-gendered instructor on faculty/student interaction, grading scale, and course organization and planning. Female students rated female instructors higher than male instructors on faculty/student interaction and grading scales. In comparison, male students rated male instructors higher than female instructors on course organization and planning.

Centra and Gaubatz (2000) found male and female students in the male-dominated discipline of natural sciences rated female instructors higher for faculty/student interaction and assignments, exams, and grading when analyzing student ratings by discipline. Centra and Gaubatz (2000) concluded that male and female instructors in natural sciences used different teaching styles, whereas female faculty used class discussion over lectures. Additionally, slight differences in the instructor ratings from the business discipline revealed female instructors rated higher for assignments, exams, grading, and course outcomes. There were no differences in the gender of instructors or students for the technology discipline.

Basow and Silberg (1987) analyzed the effects of professor gender and professor sex-typing on SET ratings. Basow and Silberg (1987) hypothesized that professor gender would interact with student gender and that male students would rate female professors lower due to professor sex typing. Over 1,000 students from a small northeastern private college rated their professors after 4 weeks of instruction. Students rated their professors on (a) scholarship, (b) organization/clarity, (c) instructor-group interaction, (d) instructor-individual interaction, (e) dynamism/enthusiasm, and (f) overall teaching ability. A

MANOVA revealed an interaction between academic discipline and SET ratings suggesting traditionally male-dominated programs (e.g., business and engineering) consistently rated female professors lower. Basow and Silberg (1987) concluded that students in male-dominated disciplines scored female professors lower than male professors because of their limited interaction with female professors and their traditional views that a college professor is a male occupation.

Narayanan et al. (2014) collected 263,492 student responses from the Dwight Look College of Engineering and the Mays Business School at Texas A&M University over seven semesters between 2007 and 2010. The SET instruments for both colleges differed with the engineering instrument rating eight areas and the business school rating 17 areas. Using ANCOVA, the results indicated that male instructors in engineering had higher SET ratings than female instructors; however, no differences were found in the SET ratings for male and female instructors in the business school. Narayanan et al. (2014) noted the percentage of female faculty in the engineering college was smaller than that of the business school and concluded that the differences found in the SET ratings of male and female faculty in engineering could be attributed to the small number of female faculty.

Gender Bias in SETs in Collegiate Aviation

Although studies examining the growth of women in collegiate aviation explored the influence female faculty had on the next generation of female aviation professionals (Ison, D., 2008; Luedtke, 1993), more needs to be known about the gender bias in SET ratings of collegiate aviation faculty. One study by Marcham et al. (2020) examined end-of-course evaluations to determine if faculty gender or faculty status influenced aviation

faculty SET ratings for online classes. Marcham et al. (2020) collected historical SET data from 683 sections of general and technical online courses taught between spring 2018 and fall 2019. Full-time and part-time faculty taught the classes using the same course materials and syllabi. Marcham et al. (2020) encouraged the faculty to provide students with their biographical information and interact with students. Marcham et al. (2020) collected SET ratings for (a) instructors' expertise of the subject matter, (b) students' overall impressions of the instructor, and (c) timeliness and quality of instructor's feedback. Marcham et al. (2020) used parametric and nonparametric tests to determine differences in SET ratings and found no significant differences between ratings based on faculty genders. Marcham et al. (2020) noted online courses might negate any gender bias that could appear in course evaluations suggesting results for courses taught in person might yield different results.

Summary

Colleges and universities commonly use SETs either alone or as part of a comprehensive evaluation system to assess the effectiveness of higher education faculty (Arreolla, 2000; Centra, 1979; Seldin, 1999). Because institutions use SET ratings in making personnel decisions such as tenure, promotion, or salary, the validity of student opinions is debatable, primarily if students evaluate factors unrelated to teaching performance (Cashin, 1999; Kreitzer & Sweet-Cushman, 2021; Seldin, 1999). The literature on gender bias in student ratings is inconclusive with mixed findings. (Bachen et al., 1999; Bennett, 1982; Chamberlin & Hickey, 2001; Feldman, 1993).

Social role theory was the theoretical lens for this research. Social role theory helps explain how traditional gender behaviors establish gender stereotypes that are

socially accepted (Eagly & Wood, 2012). Men being more dominant than women typically hold occupational positions deemed more commanding than those held by women. Based on gender expectations and the inconclusive findings in the literature regarding gender bias in student ratings (Bachen et al., 1999; Chamberlin & Hickey, 2001; Feldman, 1992; MacNell et al., 2015), the purpose of this study was to determine if gender and years of teaching experience are related to the SET ratings of collegiate aviation faculty at AABI affiliated colleges and universities. SET ratings for the overall impression of course instructors were analyzed along with four constructs related to the gender role expectations for expressiveness, immediacy, professionalism, and openness (Eagly & Wood, 2012; Feldman, 1993; Rubin, 1980; Stewart & Barraclough, 1992; Violanti et al., 2018) to determine how student ratings differ for male and female faculty.

CHAPTER 3

METHODOLOGY

The purpose of this quantitative study was to determine if gender and years of teaching experience were related to SET ratings for collegiate aviation faculty. These constructs were related to gender norms and stereotypes, including expressiveness, immediacy, professionalism, openness, and the overall SET rating. This chapter describes the methods and procedures used to conduct the study and the instruments and techniques used to collect and analyze data.

Research Questions

In this study, the following research questions were:

RQ1: What difference, if any, exists between student ratings of expressiveness for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student expressiveness ratings for female faculty and male faculty in collegiate aviation programs.

RQ2: What difference, if any, exists between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs.

RQ3: What difference, if any, exists between student ratings of professionalism for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of professionalism for male faculty and female faculty in collegiate aviation programs.

RQ4: What difference, if any, exists between student ratings of openness for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of openness for male faculty and female faculty in collegiate aviation programs.

RQ5: What difference, if any, exists between the overall student ratings of female faculty compared to that of male faculty in collegiate aviation programs?

H_0: There is no difference between the overall student ratings of female faculty and male faculty in collegiate aviation programs.

Population

The population generalized in this study included all collegiate aviation faculty employed at AABI-affiliated institutions who hold the positions of adjunct professor, lecturer, instructor, assistant professor, associate professor, or full professor.

Study Sample

The participants for this study were aviation faculty members employed at higher education institutions affiliated with AABI who provided instruction of aviation courses during the 2017-2018, 2018-2019, or 2019-2020 academic years. As of 2020, there were 38 institutions affiliated with AABI (AABI, 2020). Among the nine participating institutions in this study, males made up 77% of the faculty and females about 23%, like

the national proportion (Lutte, 2021). The number of female aviation faculty employed by the participating institutions was higher than the reported 10% found by D. C. Ison (2008) among 60 institutions 13 years ago. This study will examine student evaluations of teaching (SETs) to determine if gender and years of teaching experience are related to student ratings.

Of the 38 institutions affiliated with AABI, faculty members from 26 institutions participated in the study. Institutions were excluded from the study if their administrators did not respond to emails, they were outside the U.S., or they did not have female aviation faculty. A total of 148 faculty consisting of 63 females and 85 males were recruited to participate. Twenty-two females and 36 males who held the rankings of adjunct, lecturer, instructor, assistant professor, associate professor, or full professor submitted data for the study. The participants represented nine institutions of varying sizes and locations. This sample size was adequate for a small to medium effect size (Cohen, 1992). Table 1 describes the number of males and females represented by institution location and size.

Table 1

Distribution of Faculty by Participant Gender and Institution

Institution	Female	Male	Total	Location	Size
A	1	2	3	Northeast	Medium
B	4	13	17	Southeast	Large
C	1	0	1	Southeast	Small
D	8	16	24	Midwest	Medium
E	2	5	7	Southeast	Medium
F	2	0	2	Midwest	Medium
G	2	0	2	Southeast	Large
H	1	0	1	Southeast	Medium
I	1	0	1	Southeast	Medium

Study Solicitation

Data were collected from participants using purposeful sampling. Initially, department chairs and individual faculty members were contacted to help distribute information to the faculty at their institutions. Some institutional data was public, making it easier to collect. Where data were not publicly available, or support was not provided through contacts, the current researcher initiated direct contact with aviation faculty. Once adequate male participation was achieved, female faculty were contacted until a sufficient number was reached.

Each participant opened a link to an electronic questionnaire (Appendix A). At the bottom of the electronic questionnaire, a Dropbox file led participants to a folder where all end-of-course evaluations could be uploaded for the 2017-2018, 2018-2019, and 2019-2020 academic years. Due to the security measures created in the Dropbox folder, participants could only see their uploaded files.

The current researcher collected data, and each participating school was assigned a code. The key designating the code with the school name was kept in a separate location in a locked box to aid in de-identifying each school. Simultaneously, the current researcher de-identified each SET evaluation form of the school and participant name and assigned a number. No SET forms included students' identifying information.

Research Design

This was a quantitative study to determine if gender and years of teaching experience were related to student evaluations of teaching (SET) ratings for collegiate aviation faculty. These constructs are related to gender norms and stereotypes, including expressiveness, immediacy, professionalism, openness, and the overall SET rating. The teaching attributes analyzed were based on the gender stereotypes associated with each. Expressiveness and immediacy were considered female attributes, while professionalism and openness were deemed to be male attributes (Basow & Silberg, 1987; Bennett, 1982; Eagly & Wood, 2012; Feldman, 1992; Rubin, 1980). A fifth variable regarding the students' overall SET ratings was used to investigate comparisons related to gender and years of teaching experience among participants. SET ratings of participants were classified based on two independent variables of faculty gender (male, female) and years of teaching experience of faculty (0-6 years, 7+ years). Experience levels were included as part of the analysis to strengthen the construct validity of the study; any differences found in the study should be attributed to gender and not level of experience. The dependent variables consisted of four constructs taken from the SET ratings and consisted of expressiveness, immediacy, professionalism, and openness. An overall SET rating for aviation faculty was also included as a dependent variable.

To eliminate the extraneous variable of experience, a matched pairs design aided in grouping participants by gender and years of teaching experience (Lodico et al., 2010). The groupings for the study compared years of teaching experience of female faculty and male faculty for each dependent variable of teaching expressiveness, immediacy, professionalism, openness, and overall SET rating. The Pearson correlation between tenure status and years of teaching experience showed the two were relatively strongly correlated (R=.631), so only the number of years teaching was used in this study. Two levels of teaching experience separated novice faculty from experienced faculty. Novice faculty held 0-6 years of teaching experience, while experienced faculty held 7+ years of teaching experience (Iglesias-Martinez et al., 2014).

Data and Analysis

Data from two sources consisted of end-of-course SET evaluations completed by students for each participating aviation faculty member and completed faculty questionnaires intended to gather categorical data for each participant. SET ratings of participants were classified based on the two independent variables of gender and years of teaching experience. Participants could refuse to provide gender information or select non-traditional gender identities, but these were too few to include in this study.

Data Collection

Participants who taught aviation courses affiliated with AABI during 2017-2018, 2018-2019, and 2019-2020 provided SET data to be analyzed. Participant data consisted of a SET rating form and a questionnaire used to collect information regarding a) gender, b) institutional rank (e.g., adjunct, assistant professor, and associate professor), c) years

of experience instructing, d) tenure status, and e) primary area of instruction. The faculty questionnaire is in Appendix A.

To ensure adequate participation of both genders, faculty participated from institutions where male and female aviation faculty were employed. Of the 38 AABI-affiliated institutions, 26 employed at least one female faculty member. A total of 148 aviation faculty consisting of 63 females and 85 males were recruited to participate. Of those contacted, 22 females and 36 males participated.

Of the 58 participants, two could not provide SET statements for openness and expressiveness because their SET evaluations did not include items related to these constructs. The mean score of the gender and years of experience grouping provided data for these participants (Mertler & Vannatta, 2013). Additionally, four participants from the study were excluded due to being identified as outliers by SPSS; three were denoted as outliers (4, 38, and 52), and one as an extreme outlier (16). For immediacy, there were four cases marked as outliers (4, 15, 16, and 21). For professionalism, there were two cases designated as outliers (16 and 52). For openness, two outliers were defined (16 and 38). Two cases (6 and 52) were considered outliers for overall ratings. Data for cases 4, 16, 38, and 52 were omitted from the dataset. An outlier was represented in two or more dependent variable areas. Figure 1 illustrates the box plot identifying outliers for expressiveness. The final number of participants used in this study was 54.

A questionnaire aided in categorizing data. Each participant answered questions about his/her gender, years of experience, tenure status, and aviation specialty. The questionnaire served as a necessary tool to appropriately pair match participants into the required groupings based on gender and years of teaching experience.

Figure 1

Outliers Based on Gender for Expressiveness

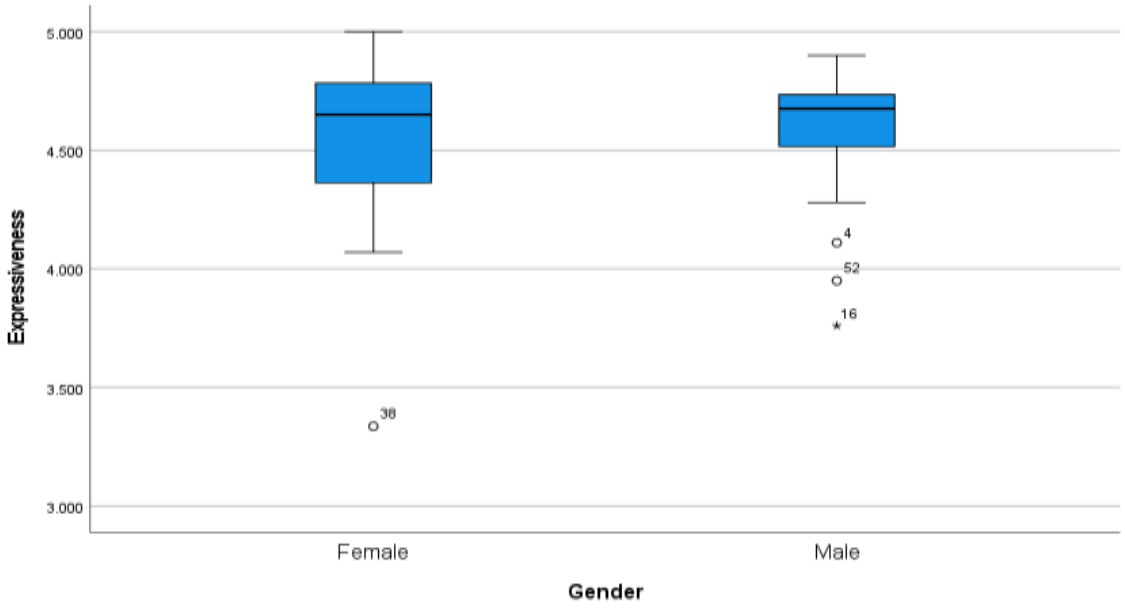


Figure 2 illustrates the box plot identifying outliers for immediacy.

Figure 2

Outliers Based on Gender for Immediacy

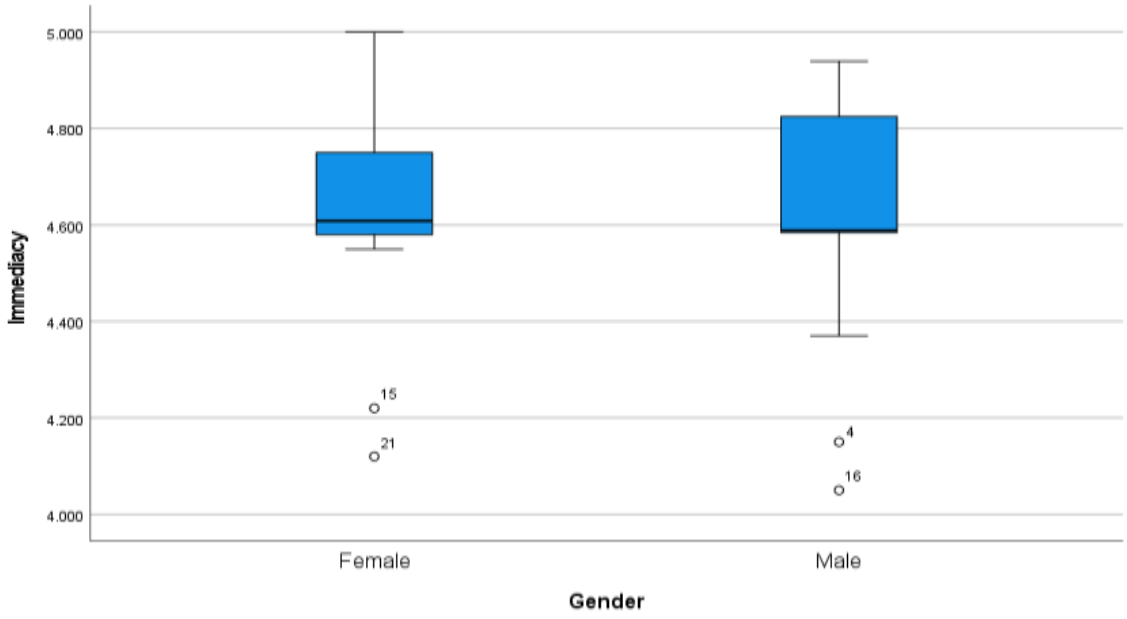


Figure 3 illustrates the box plot identifying outliers for professionalism.

Figure 3

Outliers Based on Gender for Professionalism

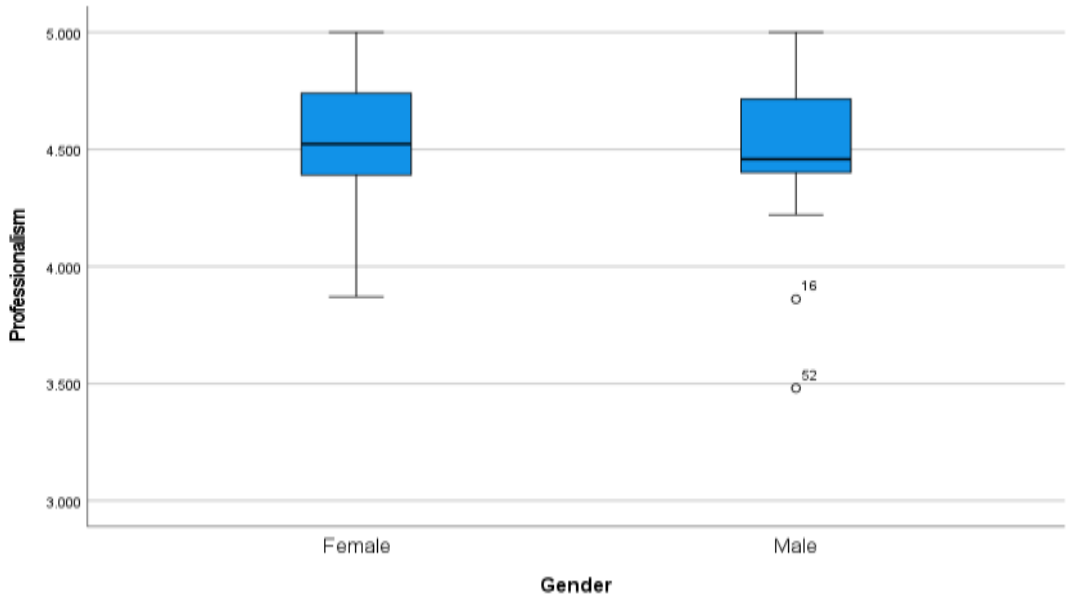


Figure 4 illustrates the box plot identifying outliers for openness.

Figure 4

Outliers Based on Gender for Openness

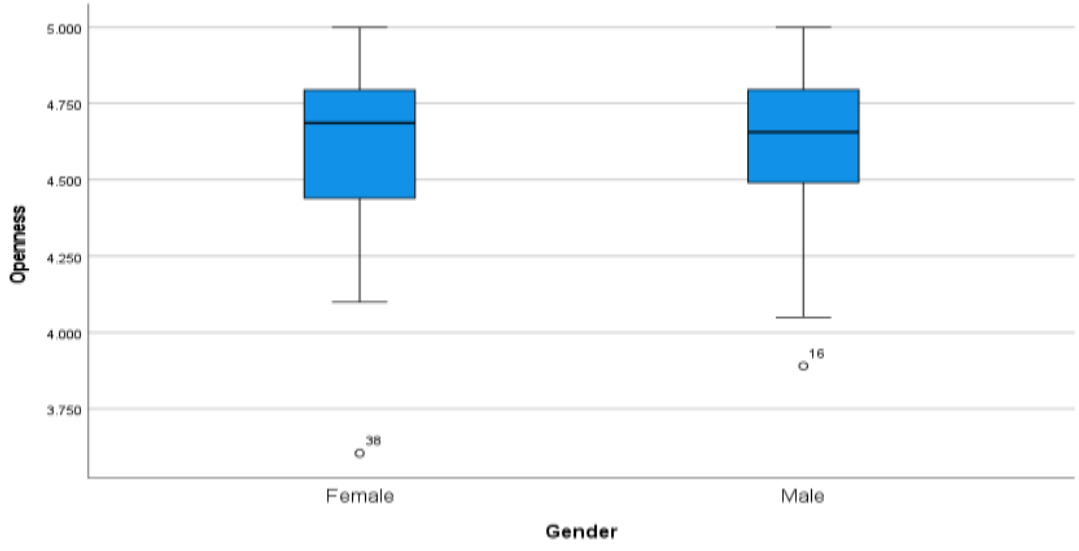
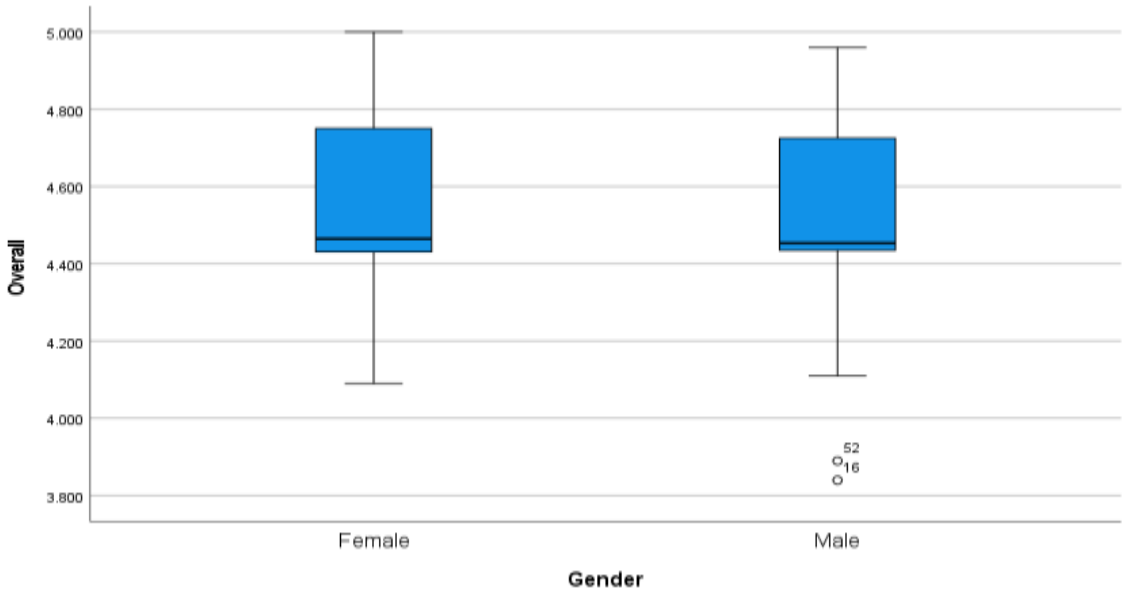


Figure 5 illustrates the box plot identifying outliers for overall.

Figure 5

Outliers Based on Gender for Overall



Data Analysis

The statistical software program SPSS version 27 aided in analyzing the data collected for this study. Categorical data was used to pair match faculty based on gender and years of teaching experience. The null hypothesis was tested at α=.05 level with equal vectors of means on the multiple dependent variables.

A factorial MANOVA aided in analyzing the student ratings for each participant grouping. A factorial MANOVA determines the differences between groups when multiple independent variables and dependent variables are present (Mertler & Vannatta, 2013). A factorial MANOVA allowed comparisons between participant groupings (independent variables) to determine the interaction between gender and years of teaching experience. A MANOVA design was appropriate for data analysis in this study

because it not only decreases the probability of making Type I errors, but it also increases the statistical power of the results by improving the chances of determining what changes occur due to the independent variable characteristics (Mertler & Vannatta, 2013). ANOVA aided in analyzing the differences between participant groupings (independent variables) and the overall rating of faculty by students (dependent variable). The ANOVA determines if a difference between means exists when two independent variables exist while simultaneously evaluating interactions between the independent variables (Mertler & Vannatta, 2013).

Instrumentation

After each academic course, students from the participating institutions used Likert-scale rating instruments to evaluate the teaching performance of aviation faculty. The evaluations were distributed either in-class or online near the end of the course. For this study, participants provided existing end-of-course evaluations from the 2017-2018, 2018-2019, and 2019-2020 academic years. Evaluations that rated the faculty members on at least one of the dependent variables of expressiveness, immediacy, professionalism, and openness were included in the study.

Because each participating institution used a different evaluation instrument, a rubric based on supporting literature aided the analysis by adding clarity and tying the study's constructs to examples from each institution's evaluation form. Appendix C illustrates each teaching construct with supporting literature across all participating institutions. After collecting the SET surveys, the current researcher classified each item on the surveys according to the definitions in Appendix C. Appendix C lists each construct used in this study along with criteria and SET statements. Once each SET

survey statement was matched with one of the research constructs, a faculty member reviewed the statements and constructs to ensure reliability. The classifications were reliable if 80% of the items were agreed upon. While reviewing the statements, statements were classified into their categories until a 96% agreement was achieved. Most institutions used a Likert-scale 1-5 (n=6) on their evaluation instruments. Because one institution used a Likert Scale 1-6 (n=1) and the remainder used 1-4 (n=2), a scale of 1-5, where 1=Strongly Disagree and 5=Strongly Agree, was applied. The surveys with a 1-4 Likert scale were recoded using a linear transformation formula to expand the range to match the 1-5 scale. One institution with a 1-6 Likert scale was recoded to reduce the range to match the 1-5 scale. When more than one statement fits a construct, an average of all applicable statements was used. For the overall ratings provided by students on the SET evaluations, an average of all statements was calculated. SPSS version 27 aided in analyzing data.

Assessing Content Validity

The SET questions of expressiveness, immediacy, professionalism, and openness were unique to each participating institution. Some institutions compiled their ratings into the relevant constructs and reported only the composite scores for each. For other institutions, the current researcher conducted a qualitative theoretical content validity assessment (Haynes et al., 1995) where SET statements were categorized based on keywords and definitions presented in the literature. When SET statements did not conform to the categorical definitions, they were removed from the analysis. Additionally, a colleague categorized the items, and an agreement of 96% was achieved,

so no further assessment was required. Further quantitative analysis of validity could not be conducted because of the differences in how institutions reported their SET data.

Threats to Validity

Inadequate Explication of Constructs

The current researcher analyzed each statement from SETs to determine if it met one of the study's four constructs (expressiveness, immediacy, professionalism, and openness). When SET statements included the wording of a construct as described in Appendix C, the rating scores for those statements were included in the study. A colleague classified the SET statements without specific wording matching the construct criteria until a 96% agreement was achieved. Statements that did not meet the construct criteria were not used.

Low Statistical Power

Because females make up less than 32% of all aviation positions in the industry (Lutte, 2021), there was concern about obtaining an appropriate number of female participants. Once an adequate number of male participants was achieved, female participants were contacted until a sufficient number of female participants was reached.

Self-Reporting

Data consisted of all SETs completed by students for the 2017-2018, 2018-2019, and 2019-2020 academic years. The current researcher understood that some participants would only send favorable student ratings. All SETs collected were used regardless of the good or unfavorable ratings.

Selection

Understanding that different institutions utilize different SET instruments when calculating student ratings of teaching effectiveness, the threat of selecting too few institutions could impact the study's outcome. To reduce the threat based on selection, institutions of various sizes and from different parts of the country were included in the study.

CHAPTER 4

RESULTS

Student evaluations of teaching (SET) are a standard evaluation tool employed in higher education institutions to measure instructor performance (Algozzine et al., 2004; Dev & Qayyum, 2017; Galbraith et al., 2011; Wachtel, 1998). In many cases, SET ratings are utilized to make personnel decisions related to promotion and tenure (Arreolla, 2000; Cashin, 1999; Centra, 2003). The weight SET ratings can hold in faculty advancement makes potential gender bias in these ratings problematic, primarily when students evaluate factors unrelated to faculty performance (Kreitzer & Sweet-Cushman, 2021).

The purpose of this quantitative study was to determine if gender and years of teaching experience are related to SET ratings for collegiate aviation faculty. These constructs are related to gender norms and stereotypes, including expressiveness, immediacy, professionalism, openness, and the overall SET rating.

Gender (male, female) and years of teaching experience (0-6, 7+) were the two independent variables in this study. The dependent variables were four constructs taken from the SET ratings consisting of expressiveness, immediacy, professionalism, and openness. The overall SET rating of faculty served as a fifth dependent variable. A factorial MANOVA at α=.05 level of significance helped analyze the student ratings for each participant grouping. Additionally, factorial ANOVA at α=.05 level of significance helped analyze the overall student ratings for each participant grouping.

The following research questions and null hypotheses addressed the multivariate main effects for each independent variable and possible interactions between factors (i.e., independent variables). Both independent variables are categorical and include gender and years of teaching experience. Two categories of years of teaching experience established novice teachers as teaching between 0-6 years and experienced teachers as teaching 7+ years. The novice category represented junior faculty members and the experienced designated senior faculty members. The dependent variables were student ratings on four teaching constructs: expressiveness, immediacy, professionalism, and openness.

RQ1: What difference, if any, exists between student ratings of expressiveness for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student expressiveness ratings for female faculty and male faculty in collegiate aviation programs.

RQ2: What difference, if any, exists between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs.

RQ3: What difference, if any, exists between student ratings of professionalism for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of professionalism for male faculty and female faculty in collegiate aviation programs.

RQ4: What difference, if any, exists between student ratings of openness for female faculty and male faculty in collegiate aviation programs?

H_0: There is no difference between student ratings of openness for male faculty and female faculty in collegiate aviation programs.

Using the SET forms from each participating institution for academic years 2017-2018, 2018-2019, and 2019-2020, a factorial MANOVA helped evaluate differences between the ratings on the teaching attributes for male and female aviation faculty in the two experience groupings. The factorial MANOVA allowed me to determine if gender and years of experience significantly affected the SET ratings related to the four teaching attributes listed.

A univariate factorial ANOVA helped evaluate differences between the SET overall rating for male and female aviation faculty and years of teaching experience. The following research question and null hypothesis address the univariate main effects for each independent variable and possible interactions between factors. The dependent variable is the overall rating calculated from each SET submitted. The factorial ANOVA allowed me to determine if the gender of the participating aviation faculty had any significant effect on the overall SET ratings.

RQ5: What difference, if any, exists between the overall student ratings of female faculty compared to that of male faculty in collegiate aviation programs?

H_0: There is no difference between the overall student ratings of female faculty and male faculty in collegiate aviation programs.

Chapter 4 provides a summary and analysis of the data. To support the data represented in this study, tables and figures illustrate the descriptive and statistical findings.

Descriptive Statistics

This quantitative comparative study consisted of a sample of 21 female aviation faculty and 33 male aviation faculty paired based on years of teaching experience. The years of experience for each gender fell into one of two categories (0-6 years and 7+ years). Table 2 shows the breakdown of participants (i.e., instructors) by age group. Table 2 presents the male and female participants for each teaching experience group. Based on the data in the table, there were more faculty participants with 7+ years of teaching experience than any other experience group.

Table 2

Descriptive Statistics: Distribution of Participant Groups

Experience	Gender	Number of Participants
0-6	Female	7
	Male	10
7+	Female	14
	Male	23
	Total	54

The dependent variable scores indicated a restriction of range in that the mean scores and standard deviations were similar. Therefore, it is unlikely to see a difference in ratings based on gender or years of experience. Table 3 illustrates the descriptive statistics for the entire sample.

Table 3

Descriptive Statistics Full Sample

Variable	N	M	SD
Expressiveness	54	4.64652	.195419
Immediacy	54	4.66907	.168806
Professionalism	54	4.56702	.250323
Openness	54	4.64885	.205193
Overall	54	4.58550	.208308

Note. N=number of participants; M=mean score; SD=standard deviation

Table 4 displays the descriptive statistics by gender and then by years of experience.

Table 4

Descriptive Statistics by Gender

Gender	Variable	N	M	SD
F	Expressiveness	21	4.61600	.249745
	Immediacy	21	4.63933	.200108
	Professionalism	21	4.54219	.316521
	Openness	21	4.63552	.226930
	Overall	21	4.57514	.252144
M	Expressiveness	33	4.66594	.152535
	Immediacy	33	4.68800	.145667
	Professionalism	33	4.58282	.201260
	Openness	33	4.65733	.193283
	Overall	33	4.59209	.178936

Note. N=number of participants; M=mean score; SD=standard deviation

Table 5 displays the descriptive statistics by gender and then by years of experience.

Table 5

Descriptive Statistics by Years of Experience

Years Total	Variable	N	M	SD
0-6	Expressiveness	17	4.63047	.229814
	Immediacy	17	4.66953	.231527
	Professionalism	17	4.55982	.342816
	Openness	17	4.67676	.204781
	Overall	17	4.59088	.254896
7+	Expressiveness	37	4.65389	.180476
	Immediacy	37	4.66886	.134637
	Professionalism	37	4.57037	.199960
	Openness	37	4.63603	.206910
	Overall	37	4.58303	.187048

Note. N=number of participants; M=means score; SD=standard deviation

Finally, Table 6 shows the mean and standard deviation scores for the four groupings.

Table 6

Descriptive Statistics by Gender and Years of Experience

Gender	Years	Variable	N	M	SD
Female	0-6	Expressiveness	7	4.54371	.324284
		Immediacy	7	4.57329	.313898
		Professionalism	7	4.39900	.451719
		Openness	7	4.63957	.290680
		Overall	7	4.49714	.346637
	7+	Expressiveness	14	4.65214	.207850
		Immediacy	14	4.67236	.112272
		Professionalism	14	4.61379	.208313
		Openness	14	4.63350	.200539
		Overall	14	4.61414	.193492
Male	0-6	Expressiveness	10	4.69120	.117615
		Immediacy	10	4.73690	.131763
		Professionalism	10	4.67240	.196717
		Openness	10	4.70280	.128033
		Overall	10	4.65650	.154214
	7+	Expressiveness	23	4.65496	.166636
		Immediacy	23	4.66674	.149003
		Professionalism	23	4.54387	.194557
		Openness	23	4.63757	.215140
		Overall	23	4.56409	.184762

Note. N=number of participants; M=mean score; SD=standard deviation.

Data Analysis Procedures

The purpose of this quantitative study was to determine if gender and years of teaching experience were related to SET ratings for collegiate aviation faculty. These

constructs are related to gender norms and stereotypes, including expressiveness, immediacy, professionalism, openness, and the overall SET rating. Before analyzing data, all data was compiled into a Microsoft Excel spreadsheet to code and de-identify for confidentiality. All data collected were assumed to be accurate and valid.

The four dependent variables were the SET ratings for expressiveness, immediacy, professionalism, and openness. The overall SET ratings of faculty served as a fifth variable to be run as a separate test. The aviation faculty's gender and years of teaching experience were the independent variables. Data analysis allowed me to determine the effect gender and years of teaching experience had on the teaching attributes of expressiveness, immediacy, professionalism, openness, and overall.

The following research questions were investigated using the MANOVA:

RQ1: What difference, if any, exists between student ratings of expressiveness for female faculty and male faculty in collegiate aviation programs?

RQ2: What difference, if any, exists between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs?

RQ3: What difference, if any, exists between student ratings of professionalism for female faculty and male faculty in collegiate aviation programs?

RQ4: What difference, if any, exists between student ratings of openness for female faculty and male faculty in collegiate aviation programs?

The following research question was investigated using the factorial ANOVA:

RQ5: What difference, if any, exists between the overall student ratings of female faculty compared to that of male faculty in collegiate aviation programs?

SET ratings from each participant were used to answer each research question. These SET ratings were for the 2017-2018, 2018-2019, and 2019-2020 academic years, and the data from the information questionnaire aided in categorizing participants. Nine assumptions associated with the MANOVA were tested to determine the appropriateness of the statistical model. For Assumptions 1-4, visual inspections of raw data were conducted, and for Assumptions 5-9, SPSS version 27 calculated the results.

Assumptions

Assumption 1: Dependent variables should be measured at the interval or ratio level.

All dependent variables in the study (expressiveness, immediacy, professionalism, and openness) were collected as interval data from each participant or institution's SET reports. SET ratings ranged from 1=Strongly Disagree, and 5=Strongly Agree. The assumption related to scores on dependent variables being at an interval level of measurement was satisfied.

Assumption 2: The independent variable should consist of two or more categorical, independent groups.

The independent variables for this research were categorical with gender and years of teaching experience, each being measured as two distinct groups. Therefore, the two groupings were independent and met the assumption.

Assumption 3: The observations within each sample are random and independent.

The SETs utilized for each institution were specific to that institution’s goals. Further, each SET used in the study was either submitted by the individual faculty participant or by a third party in a way in which the SET ratings were independent of

each other. The SET ratings of one institution did not influence the different institutions' SET ratings; therefore, Assumption 3 was met.

Assumption 4: An a*dequate sample size is required.*

A G*Power analysis (Faul et al., 2009) helped determine the appropriate sample size for an effect size of 0.39 with an α=.05, the desired power of .80, four dependent variables, and two groups. The resulting total sample size of 54 was determined. The purposeful sample consisted of 21 female aviation faculty and 33 male aviation faculty for 54 participants. The number of total participants used in this study met the assumption.

Assumption 5: There are no univariate or multivariate outliers.

For the assessment of multivariate outliers, a box plot was computed and assessed. Mertler and Vannatta (2013) explained that SPSS denotes outliers as participants with ratings between 1.5 and 3 times the length of the box plot. Extreme outliers were participants with ratings greater than three times the length of the box plot. There were both types of outliers in this study. As described in Chapter 3, four outliers were identified and omitted from the dataset. There was an outlier in two or more dependent variable areas.

Assumption 6: There is a multivariate normality.

The Shapiro-Wilk test determined the multivariate normality associated with this assumption. The Shapiro-Wilk tested the normality of each dependent variable consisting of expressiveness, immediacy, professionalism, and openness. The Shapiro-Wilk significance for expressiveness and openness produced a significance level of p=.154 for expressiveness and p=.086 for openness, indicating no significant deviation from a

normal distribution. However, the significance levels for professionalism and immediacy were $p<.001$ indicating the data was not normally distributed (Table 7). Mertler and Vannatta (2013) recommended the data be transformed to achieve normality; however, Finch (2005) found that the Type I error rates were only slightly increased due to normality issues. Because of the relatively small chance of Type I error and the robust nature of the MANOVA, I continued with the study without performing any data transformations for professionalism and immediacy.

Table 7

Shapiro-Wilk Test of Normality

Variable	Statistic	DF	Sig
Expressiveness	.968	54	.154
Immediacy	.884	54	<.001
Professional	.906	54	<.001
Openness	.962	54	.086

Note. DF=degrees of freedom; Sig=significance level

Assumption 7: There is a linear relationship between each pair of dependent variables for all combinations of groups of the independent variables.

Scatterplots were used to test this assumption. To determine the linearity of the data, scatterplots were created for each dependent variable. A visual inspection suggested a linear relationship between the dependent and independent variables.

Assumption 8: There is a homogeneity of variance-covariance matrices.

Box's Test of Equality of Covariance Matrices was conducted to test the homogeneity of variance. Due to the relatively low power of the Box test, Huberty and Petoskey (2000) propose proceeding with the study if the statistical significance of the

Box's M test is greater than .005 (p<.005). Hahs-Vaughn (2016) recommended moving with the study unless Box's test shows a probability level of p<.0001 or smaller. Thus, although statistically significant, Box's M test at p=.016 is not a violation of this assumption given these recommendations. In analyzing Levene's test, a significance level of p=.302 for openness indicated homogeneity of variance; however, the significance levels for expressiveness p=.044, immediacy p=.008, and professionalism p=.002 indicated an absence of homogeneity of variance.

Assumption 9: There is no multicollinearity

Bivariate correlations were computed to test the assumption of multicollinearity. The dependent variables used in a MANOVA should be lower than the threshold of (0.9) (Bedre, 2021; Laerd Statistics, n.d.). The Pearson correlation was less than (0.9) for all dependent variables, suggesting that the assumption for no multicollinearity was not violated (Table 8).

Table 8

Correlations of Dependent Variables

	Expressiveness	Immediacy	Professional	Openness
Expressiveness	1	.527	.551	.780
Immediacy	.527	1	.865	.601
Professional	.551	.865	1	.576
Openness	.780	.601	.576	1

Even though the data set did not meet all nine assumptions of the two-factor MANOVA, the data were sufficiently reliable to continue with the analysis to answer the research questions and hypotheses.

MANOVA Results

The current researcher conducted a two-way MANOVA with the two independent variables of gender and years of teaching experience and four dependent variables of expressiveness, immediacy, professionalism, and openness. The two-way MANOVA was appropriate based on the number of dependent and independent variables (Mertler & Vannatta, 2013). An examination of Box's Test aided in determining the homogeneity of variance-covariance. Mertler and Vannatta (2013) explained that the uniformity of variance could interfere with the MANOVA interpretation and should be tested first. The Box's Test showed that equal variances could be assumed, $F(30, 2174)=1.637$, $p=.016$, and indicated that Wilk's Lambda statistic should be used. Findings from the two-way MANOVA suggested there was no significant difference between ratings of the aviation faculty regardless of gender or years of teaching experience, Wilks' $\Lambda=.860$, $F(4, 47)=1.908$, $p=.125$, multivariate $\eta 2=.140$ (Table 9).

Table 9

Multivariate Tests for Group

Tests	Value	F	Hypothesis DF	Error DF	Sig	PES
Pillar's Trace	.140	1.908	4.000	47.000	.125	.140
Wilk's Lambda	.860	1.908	4.000	47.000	.125	.140
Hoteling's Trace	.162	1.908	4.000	47.000	.125	.140
Roy's Largest Root	.162	1.908	4.000	47.000	.125	.140

Note. F=obtained value of the F-statistic; Hypothesis DF=degrees of freedom; Error DF=degrees of freedom for the error term; Sig=significance level; PES=partial η2 or a measurement of effect size.

A follow-up analysis using factorial ANOVA with an alpha level of .05 (α=.05) was used to determine differences in each dependent variable (expressiveness, immediacy, professionalism, and openness) based on gender and experience level. A graphical representation illustrated the interaction between gender and years of teaching experience in SPSS (Figures 6, 7, 8, 9). The overlapping lines indicated an interaction between factors was present.

Results of the between-subjects effect indicated no significant difference in ratings on expressiveness, immediacy, professionalism, and openness of aviation faculty based on gender (Table 10).

Table 10

Tests Between Dependent Variables Based on Gender

DV	F	DF	Sig	PES
Expressiveness	1.638	1	.207	.032
Immediacy	2.492	1	.121	.047
Professional	1.949	1	.169	.038
Openness	.288	1	.594	.006

Note. DV=dependent variables; F=obtained value of the F-statistic; DF=degrees of freedom; Sig=significance level; PES=partial η2 or a measurement of effect size.

Results of the between-subjects effect indicated no significant difference in ratings on expressiveness, immediacy, professionalism, and openness of aviation faculty based on years of teaching experience (Table 11).

Table 11

Tests Between Dependent Variables Based on Years of Teaching Experience

DV	F	DF	Sig	PES
Expressiveness	.378	1	.542	.007
Immediacy	.083	1	.774	.002
Professional	.350	1	.557	.007
Openness	.324	1	.572	.006

Note. DV=dependent variables; F=obtained value of the F-statistic; DF=degrees of freedom; Sig=significance level; PES=partial η2 or a measurement of effect size.

However, the between-subjects results indicated a significant difference in ratings among faculty based on gender and years of teaching experience for professionalism, $F(1, 50)=5.547$, DF=1, $p=.022$, partial $\eta 2=.100$ (Table12).

Table 12

Tests Between Dependent Variables Based on Gender and Years of Teaching Experience

DV	F	DF	Sig	PES
Expressiveness	1.517	1	.224	.029
Immediacy	2.859	1	.097	.054
Professionalism	5.547	1	.022*	.100
Openness	.223	1	.639	.004

Note. DV=dependent variables; F=obtained value of the F-statistic; DF=degrees of freedom; Sig=significance level; PES=partial η2 or a measurement of effect size; *= significance at p<.05 level.

A Bonferroni post hoc analysis revealed the ratings for professionalism were not significantly different based on gender and years of teaching experience. Table13 presents the post hoc results.

Table 13

Multiple Comparisons by Group and Dependent Variable Professional

						95% confidence level for the difference	
	Group	Group	MD	SE	Sig	Lower Bound	Upper Bound
Female 0-6	1	2	-.21479	.112811	.376	-.52471	.09514
		3	-.27340	.120096	.163	-.60334	.05654
		4	-.14487	.105197	1.000	-.43388	.14414
Female 7+	2	1	.21479	.112811	.376	-.09514	.52471
		3	-.05861	.100901	1.000	-.33582	.21859
		4	.06992	.082609	1.000	-.15704	.29687
Male 0-6	3	1	.27340	.120096	.163	-.05654	.60334
		2	.05861	.100901	1.000	-.21859	.33582
		4	.12853	.092310	1.000	-.12507	.38213
Male 7+	4	1	.14487	.105197	1.000	-.14414	.43388
		2	-.06992	.082609	1.000	-.29687	.15704
		3	-.12853	.092310	1.000	-.38213	.12507

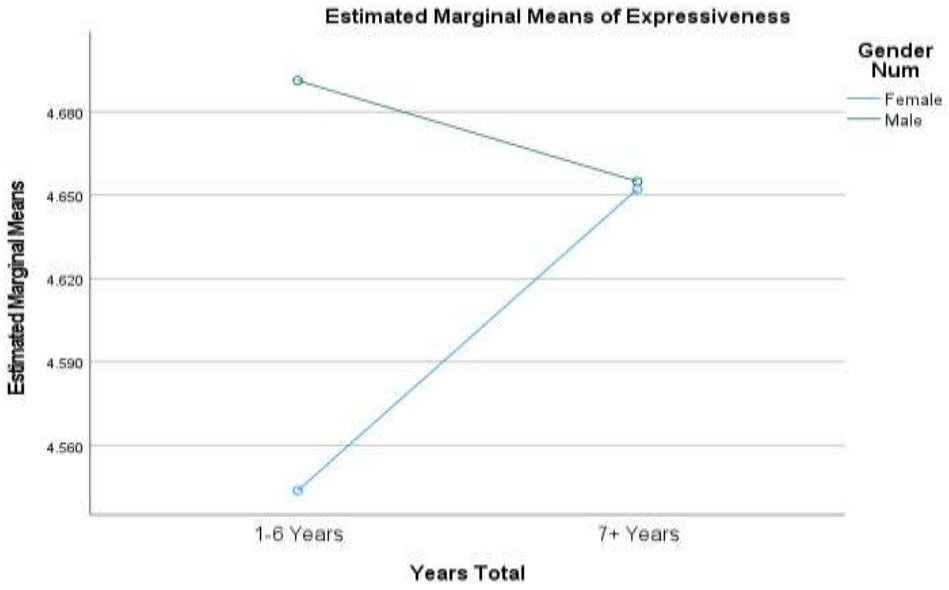
Note. MD=mean difference, SE=standard error, Sig=significance or probability level.

Research Question 1 - Expressiveness

To answer Research Question 1, a MANOVA generated findings showing SET ratings for the expressiveness of female and male aviation faculty in collegiate aviation programs were not significantly different regardless of years of experience. As Figure 6 demonstrates, a graphical representation of the data suggested there might not be an interaction between years of experience and gender.

Figure 6

Interaction Between Gender and Years of Teaching Experience for Expressiveness



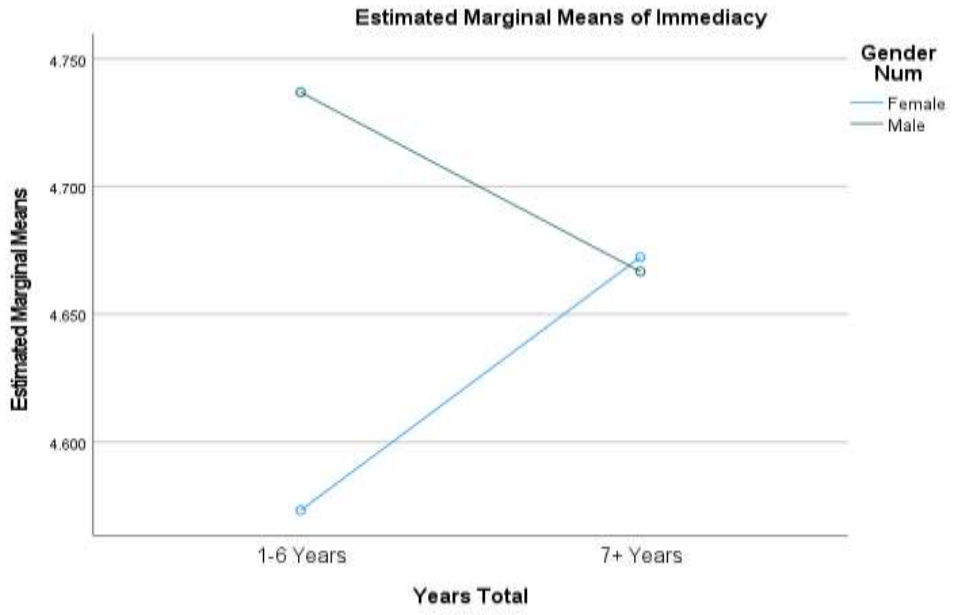
The null hypothesis stated no difference between student expressiveness ratings for female faculty and male faculty in collegiate aviation programs. There was no statistically significant difference between female aviation faculty (M=4.61600, SD=.249745) and male aviation faculty (M=4.66594, SD=.152535). Therefore, I failed to reject the null hypothesis.

Research Question 2 - Immediacy

To answer Research Question 2, a MANOVA generated findings showing SET ratings for immediacy for female and male aviation faculty in collegiate aviation programs were not significantly different regardless of years of experience. As Figure 7 demonstrates, a graphical representation of the data suggested there might not be an interaction between years of experience and gender.

Figure 7

Interaction Between Gender and Years of Teaching Experience for Immediacy



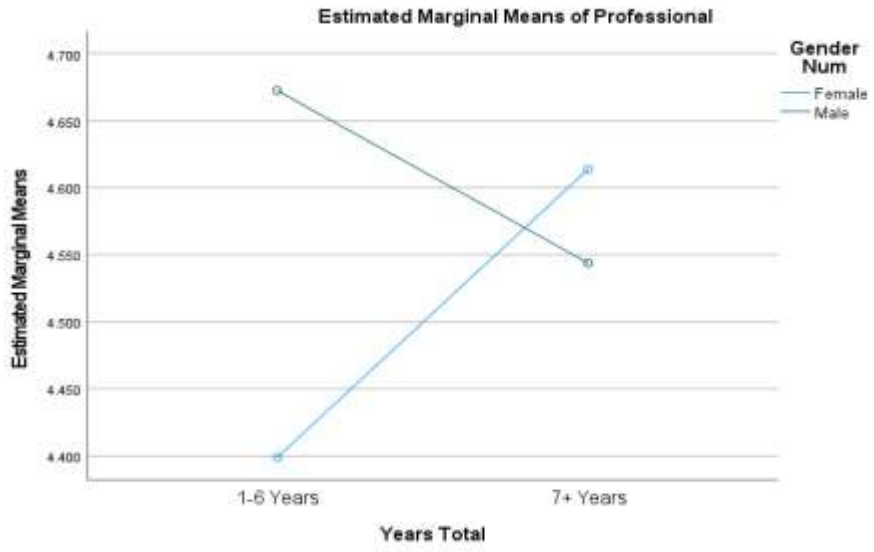
The null hypothesis stated no difference between student ratings of immediacy for female faculty and male faculty in collegiate aviation programs. There was no statistically significant difference between student ratings of immediacy for female aviation faculty (M=4.63933, SD=.200108) and male aviation faculty (M=4.68800, SD=.145667). Therefore, I failed to reject the null hypothesis.

Research Question 3 - Professionalism

To answer Research Question 3, a MANOVA generated findings indicating the SET ratings for professionalism for female and male aviation faculty in collegiate aviation programs were significantly different $F(1, 50)=5.547$, $p=.022$, partial $\eta 2=1.00$, which was also suggested in the graphical representation (Figure 8).

Figure 8

Interaction Between Gender and Years of Teaching Experience for Professionalism



The Bonferroni post hoc analysis revealed that novice female aviation faculty's SET ratings were not significantly different from experienced female aviation faculty (p=.376). The SET ratings of novice female aviation faculty were not significantly different from those of novice male aviation faculty (p=.163) or experienced male aviation faculty (p=1.000).

The null hypothesis stated no difference between student ratings of professionalism for male faculty and female faculty in collegiate aviation programs. There was no statistically significant difference between student ratings of professionalism for female aviation faculty (M=4.54219, SD=.316521) and male aviation faculty (M=4.58282, SD=.201260). Therefore, I failed to reject the null hypothesis.

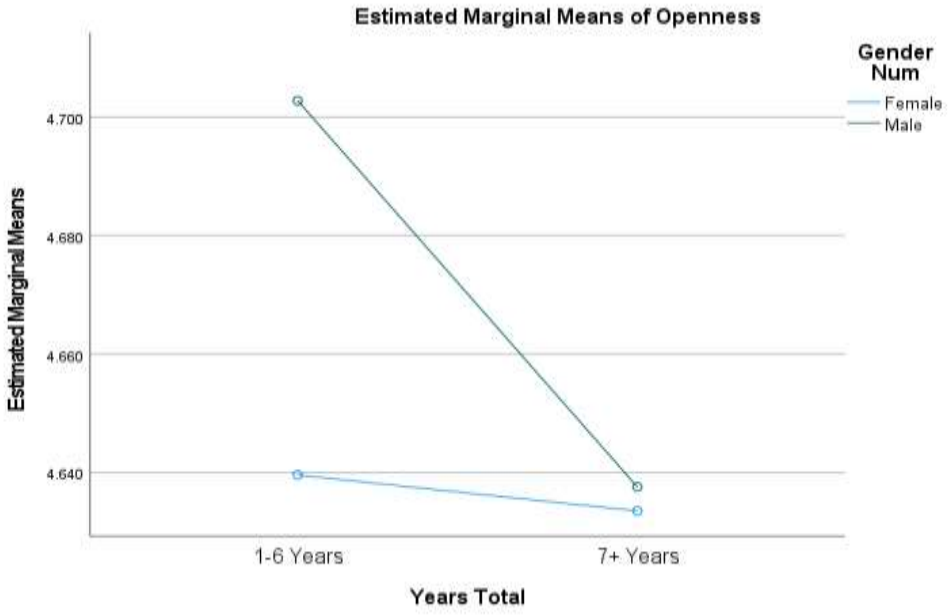
Research Question 4 - Openness

To answer Research Question 4, a MANOVA generated findings showing SET ratings for openness for female and male aviation faculty in collegiate aviation programs

were not significantly different regardless of years of experience. The null hypothesis stated no difference between student ratings of openness for male faculty and female faculty in collegiate aviation programs. A graphical representation of the data suggested there might not be an interaction between years of experience and gender (Figure 9).

Figure 9

Interaction Between Gender and Years of Teaching Experience for Openness



There was no statistically significant difference between student ratings of openness for female aviation faculty (M=4.63552, SD=.226930) and male aviation faculty (M=4.65733, SD=.193283). Therefore, I failed to reject the null hypothesis.

ANOVA Results

Research Question 5 – Overall Ratings

To answer Research Question 5, an ANOVA generated findings showing the overall SET ratings for female and male aviation faculty in collegiate aviation programs were not significantly different regardless of years of experience. A factorial ANOVA was appropriate for examining the differences between variables while simultaneously examining the interaction between more than one independent variable (Mertler & Vannatta, 2013). The alpha level of 0.05 identified statistically significant effects of the independent variables on the dependent variables.

SPSS version 27 was used for the complete data analysis. Three assumptions associated with the ANOVA were tested to determine the appropriateness of the statistical model.

Assumption 1: The independent variable should consist of two or more categorical, independent groups.

The independent variables for this research were categorical gender and years of teaching experience measured as two distinct groups. Therefore, the two groupings were independent and met the assumption.

Assumption 2: There is univariate normality.

The Shapiro-Wilk test aided in determining the normality associated with this assumption. The Shapiro-Wilk tested the normality of the dependent variable of overall SET ratings. The Shapiro-Wilk significance for female aviation faculty had a significance level of p=.225. However, the significance levels for male aviation faculty produced a significance level of p<.001, indicating the data was not normally distributed (Table 14).

Because the chance of error is present when normality is violated, data is recommended to be transformed to achieve normality (Mertler & Vannatta, 2013). However, Finch (2005) found that the Type I error rates slightly increased due to normality issues. Because of the relatively small chance of Type I error and the robust nature of the ANOVA, I continued with the study.

Table 14

Shapiro-Wilk Test of Normality for Overall Rating

Variable	Statistic	DF	Sig
Female	.941	21	.225
Male	.835	33	<.001

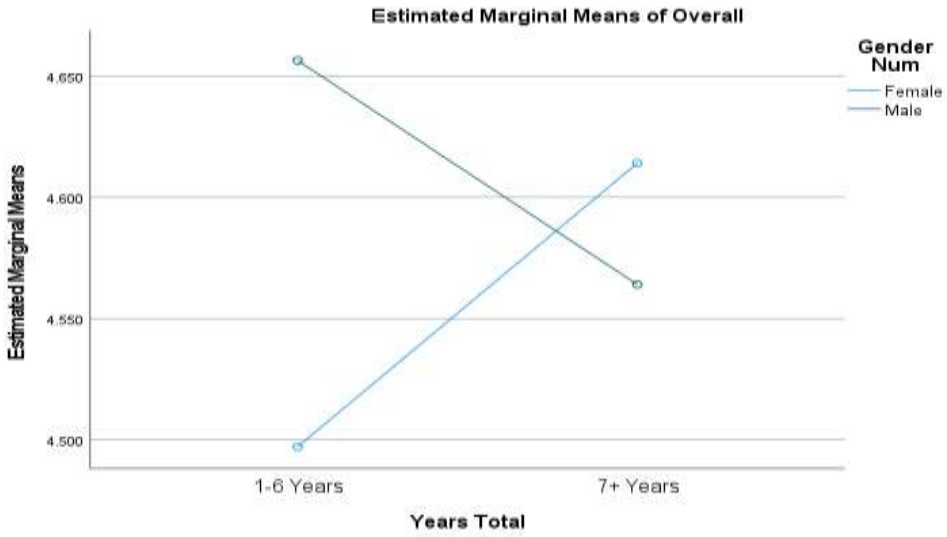
Note. DF=degrees of freedom; Sig=significance level

Assumption 3: *There is a homogeneity of variance-covariance matrices.*

Levene's test for equal variance indicated an absence of homogeneity of variance $F(3, 50)=4.038$, $p=.012$. A graphical representation of the interaction between gender and years of teaching experience was created in SPSS (Figure 10). The overlapping lines indicated a possible interaction between factors was present. Although the chart shows an interaction may be present, the interaction may not be statistically significant once the ANOVA is run (Mertler & Vannatta, 2013).

Figure 10

Visual Representation of Interaction Between Factors



The ANOVA results were presented in Table 15 and they showed no significant main effect for gender, (F(1, 50)=.768, p=.385, partial η2=.015) and years of teaching experience, (F(1, 50)=.039, p=.844, partial η2=.001). Interaction between factors was also not significant, F(1, 50)=2.821, p=.009, partial η2=.053). The null hypothesis stated no significant difference between the overall student ratings of female faculty and male faculty in collegiate aviation programs. There was no statistically significant difference between the overall student ratings for female aviation faculty (M=4.57514, SD=.055022) and male aviation faculty (M=4.59209, SD=.031149). Therefore, I failed to reject the null hypothesis.

Table 15

Two-way ANOVA Summary

Source	SS	DF	MS	F	Sig	PES
Between treatments	.127	3	.042	.975	.412	.055
Gender	.033	1	.033	.768	.385	.015
Years of Experience	.002	1	.002	.039	.844	.001
Gender x Years of experience	.123	1	.123	2.821	.099	.053
Within treatments	2.173	50	.043			
Total	1137.748	54				

Note. SS=sum of squares; DF=degrees of freedom*;* MS=mean squared; F=obtained value of the F-statistic; Sig=significance level; PES=partial η2 or a measurement of effect size.

Summary

This chapter described SETs and their roles in personnel decisions at higher education institutions. Participants consisted of a sample of 54 aviation faculty members working at AABI-affiliated institutions during the 2017-2018, 2018-2019, and 2019-2020 academic years. Descriptive statistics of the sample consisted of gender and years of teaching experience. This quantitative comparative study aimed to determine if differences in gender and years of teaching experience (independent variables) were related to the SET ratings of expressiveness, immediacy, professionalism, openness, and overall ratings (dependent variables).

Results from the statistical analysis for Research Questions 1-5 were outlined in the previous sections. As stated earlier, there were no statistically significant differences between genders on the SET of aviation faculty. These findings contradict much of the literature and will be discussed in the next chapter.

CHAPTER 5

DISCUSSION

The purpose of this quantitative study was to determine if gender and years of teaching experience related to student evaluations of teaching (SET) ratings for collegiate aviation faculty at AABI-affiliated institutions. These constructs are related to gender norms and stereotypes, including expressiveness, immediacy, professionalism, openness, and the overall SET rating.

Most higher education institutions rely on student evaluations to measure teaching effectiveness (Arreolla, 2000; Centra, 1979; Hoffman & Oreopoulos, 2009). Teaching constructs matching female and male stereotypes were identified with examples (Bachen et al., 1999; Basow, 2000; Basow & Silberg, 1987; Bennett, 1982; Kierstead et al., 1988; Rubin, 1980).

In the present study, no differences existed in the mean SET ratings of male and female collegiate aviation faculty, regardless of years of teaching experience, expressiveness, immediacy, professionalism, openness, and overall. The results of this study were similar to studies that found little to no evidence of differential ratings associated with faculty gender (Bennett, 1982; Feldman, 1993; Marcham et al., 2020); however, other research has demonstrated that differences in student perceptions of male and female faculty do exist (Bachen et al., 1999; Basow, 2000; Basow & Silberg, 1987; Centra & Gaubatz, 2000; Chamberlin & Hickey, 2001). In some instances, gender was

the sole explanation for the difference in SET ratings (Boring et al., 2016; MacNell et al., 2015; Mitchell & Martin, 2018). Other studies found students gave higher SET ratings to faculty of the same gender as themselves (Bachen et al.,1999; Basow & Silberg, 1987; Chamberlin & Hickey, 2001) or gave ratings based on gender stereotypes (Basow, 2000; Boring, 2016; Kierstead et al., 1988).

Female faculty descriptions that aligned with common stereotypes were helpfulness (Basow, 2000) and approachability (Bachen et al., 1999), while male faculty descriptions were organized (Basow, 2000) and open-minded (Rubin, 1980). In most studies, the SET ratings for female faculty may have been influenced by the students' gender role expectations. Female faculty that met the gender expectations of the students were rewarded (Kierstead et al., 1988) or penalized (Basow & Silberg, 1987; Bennett, 1982) when they did not live up to student expectations. The same rewards and punishments did not seem to apply to male faculty. For instance, Kierstead et al. (1988) found SET ratings for male faculty were not affected by the amount of interaction they had with students.

In the present study, teaching effectiveness ratings related to expressiveness and immediacy were characterized as feminine traits (Basow & Silberg, 1987; Bennett, 1982; Kierstead et al., 1988). In contrast, the teaching effectiveness ratings associated with professionalism and openness were masculine traits (Basow, 2000; Centra & Gaubatz, 2000; Chamberlin & Hickey, 2001; Eagly & Wood, 2012). The lack of significant differences in the current study suggests that students in aviation do not associate these traits with the gender of aviation faculty.

The findings of this study are significant because they add to a growing body of knowledge on the validity of SETs used for measuring teaching effectiveness in higher education. Additionally, this study contributes to research targeting aviation degree programs. Although there is existing research on the assessment processes of aviation programs (Lyons, 2021), very little research explores how aviation students evaluate the teaching effectiveness of aviation faculty. The findings of this study may assist future research in specialized programs like aviation.

Limitations

The current study aimed to determine if gender and years of teaching experience were related to SET ratings for collegiate aviation faculty at AABI-affiliated institutions. The gender of the students completing the SETs was not considered. Including and evaluating the interaction of student gender with faculty gender might have provided additional information that would help explain why no differences in SET ratings of male and female faculty were observed.

The current study was designed as a quantitative study to determine to what degree gender and years of teaching experience were related to SETs of collegiate aviation faculty members. This study could have benefited from using qualitative methods to bring out differences in gender not represented in SET evaluations used by the institutions. Interviews with students or faculty could have provided information vital for explaining the SET ratings or identifying differences not measured by SETs.

Delimitations

The present study was limited to those institutions affiliated with AABI. This only allowed data from 38 institutions. In the United States, the UAA recognizes over 100 colleges and universities with organizational memberships (UAA, 2021). Different results may be found if the study included colleges and universities associated with UAA. Including more institutions with aviation programs might provide more diversity and allow more institutions and regions to be included in the results.

Implications for Practice

This study did not find that gender or years of experience were related to the SET ratings of male and female collegiate aviation faculty for expressiveness, immediacy, professionalism, openness, or overall ratings. However, administrators should be aware of potential biases such as grades, course workload, age, and gender that influence student perceptions of teaching. Though SETs may help provide feedback for improving teaching, their use should be limited in making personnel decisions such as tenure and promotion. SETs should be used as part of an evaluation system incorporating peer and administrative evaluations, providing a more comprehensive context for assessing faculty competence.

Implications for Future Research

To help understand why no significant differences were found in mean SET ratings of aviation faculty based on gender and years of teaching experience, contradicting much of the literature (Bachen et al., 1999; Basow & Silberg, 1987; Chamberlin & Hickey, 2001; Kierstead et al., 1988; MacNell et al., 2015), the current

researcher reexamined the descriptive statistics and noticed the range of scores within each grouping. As Table 16 demonstrates, novice female faculty had a wider SET range than any other grouping. Additionally, novice female faculty had the lowest means in each category, while novice male faculty had the highest means. However, this difference was also not significant.

Table 16

Descriptive Statistics by Construct and Participant Group

	Females 0-6		Males 0-6		Females 7+		Males 7+	
Construct	M(SD)	R	M(SD)	R	M(SD)	R	M(SD)	R
Expressiveness	4.54(.32)	.930	4.69(.12)	.380	4.65(.21)	.620	4.65(.17)	.620
Immediacy	4.57(.31)	.880	4.74(.13)	.306	4.67(.11)	.303	4.67(.15)	.570
Professionalism	4.4(.45)	1.13	4.67(.20)	.508	4.61(.21)	.566	4.54(.20)	.603
Openness	4.64(.29)	.900	4.70(.13)	.393	4.63(.20)	.662	4.64(.22)	.700
Overall Rating	4.5 (.347)	.910	4.66(.15)	.389	4.61(.19)	.525	4.56(.19)	.590

Note. M=Mean; SD=Standard deviation; R=Range.

According to the MANOVA, the differences in the mean scores for the novice female aviation faculty were not due to gender and teaching experience; however, the broader ranges of the scores in each teaching construct and overall suggested an interaction with gender and years of experience. It might be possible that the (a) novice male aviation faculty, (b) experienced female aviation faculty, and (c) experienced male aviation faculty have similar teaching styles. Alternately, it may take more time for novice female aviation faculty to adapt their teaching style to that of the other groupings. It is also possible there exists more bias for novice female aviation faculty. Future studies could explore what, if anything, explains these differences.

Considering that collegiate aviation programs do not follow the traditional academic framework (Smith, 2002), the present study's findings may not be surprising. Practical aviation experience, not academic degrees, tends to be more effective for developing aviation professionals (Lindseth, 1996; Smith, 2002; Ison, D. C., 2008). By the time an aviation faculty member gets to a higher education institution, he/she has undergone extensive training either in the military or through the required Federal Aviation Administration (FAA) certification process necessary to provide aviation instruction (Ison, D. C., 2008). Even aviation management faculty must have adequate practical work experience to provide quality instruction to students (Smith, 2002). Additionally, collegiate aviation programs with credible faculty rich in professional experiences aid in attracting and retaining students. Therefore, future studies may explore the factors that explain how aviation faculty differs from other male-dominated programs.

It is also possible personality filters aid in shaping the homogeneity among faculty drawn to aviation. Given the amount of training and practical work experience achieved by most aviation faculty, it would not be surprising that a certain level of homogeneity exists among them, especially regarding personality. Personality assessments such as the Myers-Briggs Type Indicator are commonly used to identify personality traits predicting job success (Muchinsky, 2006). Aviation faculty and even students may share personality traits that attract them to aviation. The training and experience coupled with the personality type of collegiate aviation faculty might be similar to the extent that any gender differences are not evident in SET teaching ratings. Therefore, future studies may explore the personality traits of those attracted to aviation.

The current study was designed as a quantitative study to evaluate the differences between SET ratings of male and female aviation faculty members. This study could have benefited from using qualitative methods to bring out differences in gender not represented in SET evaluations used by the institutions. Interviews with students or faculty could have provided information vital for explaining the SET ratings or identifying differences not measured by SETs. Future research might benefit from a mixed methodological design.

Conclusion

Assessing faculty effectiveness using SETs is a common practice for colleges and universities (Arreolla, 2000; Centra, 1979; Seldin, 1999). Because the validity of SETs for assessing faculty performance is inconclusive (Bachen et al., 1999; Bennett, 1982; Chamberlin & Hickey, 2001; Feldman, 1993), making personnel decisions such as tenure and promotion based on student opinion is questionable, primarily if students evaluate factors unrelated to teaching performance (Cashin, 1999; Kreitzer & Sweet-Cushman, 2021; Seldin, 1999).

The purpose of this study was to determine if gender and years of teaching experience were related to the SET ratings of collegiate aviation faculty at AABI-affiliated colleges and universities. SET ratings for the overall impression of course instructors were analyzed along with four constructs related to the gender role expectations for expressiveness, immediacy, professionalism, and openness (Eagly & Wood, 2012; Feldman, 1993; Rubin, 1980; Stewart & Barraclough, 1992; Violanti et al., 2018). The expectations for expressiveness and immediacy were considered female traits, and the expectations for professionalism and openness were considered male traits. Based

on gender expectations and the inconclusive findings in the literature regarding gender bias in student ratings (Bachen et al., 1999; Chamberlin & Hickey, 2001; Feldman, 1992; MacNell et al., 2015), these constructs were studied to determine how student ratings differ for male and female collegiate aviation faculty.

In the present study, no differences existed in the mean SET ratings of male and female collegiate aviation faculty, regardless of years of teaching experience for expressiveness, immediacy, professionalism, openness, and overall ratings. However, administrators should use caution when putting too much weight on the perceptions of students, especially when they may rate faculty based on biases and not their actual teaching performance.

REFERENCES

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Ganesh, M., & Spooner, F. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching, 52*(4),134-141.

Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles, 49*(9-10), 507-516.

Arreolla, R. A. (2000). *Developing a comprehensive faculty evaluation system*. Anker Publishing Company, Inc.

Aviation Accreditation Board International. (2021). *Glossary of terms.* https://www.aabi.a/ero/about-aabi/forms-and-publications

Bachen, C. M., McLoughlin, M. M., & Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education, 48*(3), 193.

Barnett, R. C. (2014). *Role theory*. Encyclopedia of Quality of Life and Well-Being Research. https://link.springer.com/referenceworkentry/10.1007/978-94-007-0753-5_2535

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology, 79*(3), 308-314.

Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles, 43*(5-6), 407-417.

Bedre, R. (2021, October 1). MANOVA using R (with examples and code). *Data Science Blog*. https://www.reneshbedre.com/blog/manova.html#assumptions-of-manova

Bennett, S. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology, 74*(2), 170-179.

Bharadwaj, S., Futrell, C. M., & Kantak, D. M. (1993). Using student evaluations to improve learning. *Marketing Education Review, 3*(2), 16-21.

Blackhart, G. C., Peruche, M. B., DeWall, N. C., & Joiner T. E., Jr. (2006). Factors influencing teaching evaluations in higher education. *Teaching of Psychology, 33*(1), 37-39. https://doi.org/10.1207/s15328023top3301_9

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Science Open Research, 0*(0), 1-11. https://doi.org/10.14293/S2199-10061.1.SOR-EDU.AETBZC.v1

Boring, A. (2016). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27-41.

Cashin, W. E. (1999). *Student ratings of teaching: Uses and misuses in changing p**ractices in evaluating teaching*. Anker Publishing Company, Inc.

Centra, J. A. (1979). *Determining faculty effectiveness*. Jossey-Bass Inc.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*(5), 495-517.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education, 71*(1), 17-33.

Chamberlin, M. S., & Hickey, J. S. (2001). Student evaluations of faculty performance: The role of gender expectations in differential evaluations. *Educational Research Quarterly, 25*(2), 3-11.

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education, 13*(4), 321-341.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science 1*(3), 98-101. https://doi.org/10.1111/1467-8721.ep10768783

Dev, S., & Qayyum, N. (2017). Major factors affecting students' perception towards faculty evaluation of teaching. *Journal of Social Studies Education Research, 8*(3), 149-167.

Eagly, A. (1987). *Sex differences in social behavior: A social-role interpretation*. Lawrence Erlbaum Associates, Inc., Publishers.

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social Psychology* (Vol. 2) (pp. 458–476). Sage Publications. https://doi.org/10.4135/9781446249222

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Feldman, K.A. (1992). College students' views of male and female college teachers: Part I- Evidence from the social laboratory and experiments. *Research in Higher Education, 33*(3), 317-375.

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II-Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, *34*(2), 151-211.

Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology European Journal of Research Methods for the Behavioral and Social Sciences, 1*(1), 27-38.

Froehlich, L., Tsukamoto, S., Morinaga, K., Uchida, Y., Keller, M., Sturmer, S., Martiny, S. & Trommsdorff, G. (2022). Gender stereotypes and expected backlash for female STEM students in Germany and Japan. *Frontiers in Education, 6*, 1-15. https://doi.org/10.3389/feduc.2021.793486

Galbraith, C. S., Merrill, G. B., & Kline, D. M. (2011). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education*, *53*(3), 353-374.

Greenwald A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, *89*(4), 743-751.

Griffin, T. J., Hilton, III, Plummer, K., & Barret, D. (2014). Correlation between grade point averages and student evaluation of teaching scores: Taking a closer look. *Assessment & Evaluation in Higher Education*, *39*(3), 339-348.

Hahs-Vaughn, D. (2016). Applied multivariate statistical concepts. Taylor & Francis.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247.

Heilman, M. E. (1997). Sex discrimination and the affirmative action remedy: The role of sex stereotypes. *Journal of Business Ethics*, *15*(9), 877-889. https://doi.org/10.1023/Q:1017927002761

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues, 57*(4), 657.

Hornstein, H. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education, 4*(1), 1-8.

Holec, V., & Marynowski, R. (2020). Does it matter where you teach? Insights from a quasi-experimental study of student engagement in an active learning classroom. *Teaching & Learning Inquiry, 8*(2), 140-163.

Huberty, C. J., & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 183-208). Academic Press.

Iglesias-Martinez, M. J., Lozano-Cabezas, I., & Martinez-Ruiz, M. A. (2014). Listening to voices of novice lecturers in higher education: A qualitative study. *International Journal of Teaching and Learning in Higher Education, 26*(2) 170-181.

Ison, D. (2008). The status of women faculty in four-year aviation higher education programs. *Journal of Women in Educational Leadership, 6*(3), 201-210.

Ison, D. C. (2008). Pathways to the aviation professoriate: An investigation into the attributes of professional pilot education faculty. *Collegiate Aviation Review, 27*(2), 28-44.

Ison, D. C., (2010). The future of women in aviation: Trends in participation in postsecondary aviation education. *Journal of Aviation/Aerospace Education & Research, 19*(3), 27-40.

Joye, S., & Wilson, J. (2015). Professor age and gender affect student perceptions and grades. *Journal of the Scholarship of Teaching and Learning, 15*(4), 126-138.

Kierstead, D., D'agostino, P., & Dill, H. (1988). Sex role stereotyping of college professors: Bias in students' ratings of instructors. *Journal of Educational Psychology, 80*(3), 342-344.

Kreitzer, R. J., & Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics, 20*(1), 73-84. doi:10.1007/s10805-021-0900-w

Laerd Statistics. (n.d.). *Two-way MANOVA in SPSS statistics.* https://statistics.laerd.com/spss-tutorials/two-way-manova-using-spss-statistics.php

LeFebvre, L., & Allen, M. (2014). Teacher immediacy and student learning: An examination of lecture/laboratory and self-contained course sections. *Journal of the Scholarship of Teaching and Learning, 14*(2), 22-45.

Lindseth, P. D. (1996). *Identifying indicators of program quality in United States baccalaureate aviation programs* (Publication No. 9624672) [Doctoral dissertation, University of Michigan]. ProQuest Dissertations and Theses Global.

Little, W. (2016). *Introduction to sociology* (2nd ed.). BC Campus. https://opentextbc.ca/introductiontosociology2ndedition/

Lodico, M., Spauling, D., & Voegtle, K. (2010). *Methods in educational research: From theory to practice*. Jossey-Bass. http://stikespanritahusada.ac.id

Lyons, J. G. (2021). Specialized accreditation in collegiate aviation: A case study on evaluative inquiry practices required by the Aviation Accreditation Board International. *Essays in Education, 28,* 1-36. https://openriver.winona.edu/eie/vol28/iss1/1/

Luedtke, J. R. (1993). *Maximizing participation of women in collegiate aviation education* (Publication No. 9407262). [Doctoral dissertation, Wichita State University]. ProQuest Dissertations & Theses Global.

Lutte, R. (2021). *Women in aviation: A workforce report*. University of Nebraska at Omaha. https://www.researchgate.net/publication/357032586

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching? *Innovative Higher Education, 40*(4), 291-303.

Marcham, C. L., Ade, A. M., Clark, P., Marion, J. (2020). Bias and trends in student evaluations in online higher education settings. *Collegiate Aviation Review International, 38*(2), 34-50. http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8036/7417

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-388.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders*? Journal of Education Psychology, 92*(1), 202-208.

Mertler, C.A. & Vannatta, R. A. (2013). *Advanced and multivariate statistical methods: Practical application and interpretation* (5th ed.). Pyrczak Publishing.

Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *PS-Political Science and Politics, 51*(3), 548-652. https://doi.org/10.1017/S104909651800001X

Muchinsky, P. (2006). *Psychology applied to work* (8th ed.). Thomas Wadsworth.

Murray, H. (1991). Effective teaching behaviors in the college classroom. In J. Smart (Ed.), *Higher education: Handbook of theory and research.* Agathon Press.

Narayanan, A., Sawaya, W. J., & Johnson, M. D. (2014) Analysis of differences in nonteaching factors influencing student evaluation of teaching between engineering and business classrooms. *Decision Sciences Journal of Innovative Education*, *12*(3), 233-265.

Rubin, R. B. (1980). Ideal traits and terms of address for male and female college professors. *Journal of Personality and Social Psychology*, *41*, 966-97.

Schonwetter, D., Menec, V., & Perry, R. (1995, April 18-22). *An empirical comparison of two effective college teaching behaviors: Expressiveness and organization* [Paper presentation]. American Educational Research Association Annual Meeting, San Francisco, CA, United States.

Seldin, P. (1999). *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions*. Anker Pub.

Smith, D. E. (2002). The demise of collegiate aviation programs with the best of intentions. *Journal of Aviation/Aerospace Education & Research, 12*(1), 13-14.

Stewart, R. A., & Barraclough, R. A. (1992, February 21-25). *Immediacy and enthusiasm as separate dimensions of effective college teaching: A test of Lowman's model on student evaluation of instruction and course grades* [Paper presentation]. Western Speech Communication Association 63rd Annual Meeting, Boise, ID, United States.

Stonebraker, R., & Stone, G. (2015). Too old to teach? The effect of age on college and university professors. *Research in Higher Education, 56*(8), 793-812.

Thornton, B., Adams, M., & Sepehri, M. (2010). The impact of students' perceptions of course difficulty, workload, and pace on faculty evaluations. *Contemporary Issues in Education Research, 3*(12), 1-6.

University Aviation Association. (2021). *Colleges and universities.* https://www.uaa.aero/colleges_universities.php

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education, 23*(2), 191. https://doi.org/10.1080/0260293980230207

Wilson, J. H., Beyer, D., & Monteiro, H. (2014). Professor age affects student ratings: Halo effect for younger teachers. *College Teaching*, *62*(1), 20-24.

Verniers, C., & Martinot, D. (2015). Characteristics expected in fields of higher education and gender stereotypical traits related to academic success: A mirror effect. *Social Psychology of Education*, *18*(4), 719-733. https://doi.org/10.1007/s11218-015-9312-z

Violanti, M. T., Kelly, S. E., Garland, M. E., & Christen, S. (2018). Instructor clarity, humor, immediacy, and student learning: Replication and extension. *Communication Studies*, *69*(3), 251-262. https://doi.org/10.1080/10510974.2018.1466718

APPENDIX A

AVIATION FACULTY SURVEY

Aviation Faculty Survey

Thank you for your participation in this research project. Please complete the following information to be used for classification purposes.

Participant Initials _____

I attest that I have read and understood the following description of this study and its purposes and methods. I understand that my participation in this research is strictly voluntary and my participation or refusal to participate in this study will not affect my relationship with Louisiana Tech University. Further, I understand that I may withdraw at any time or refuse to answer any questions without penalty. Upon completion of the study, I understand that the results will be freely available to me upon request. I understand that the results of the material will be confidential, accessible only to the principal investigators, myself, or a legally appointed representative. I have not been requested to waive nor do I waive any of my rights related to participating in this study.

____ I do ____ I do not

1. **What is our gender identity?**
   _____ Female _____ Male _____non-binary

   _____Other _____ Prefer not to say

2. **What is your current institutional position (choose one)?**

   _____ Academic Instructor _____ Lecturer _____ Adjunct Professor

   _____ Assistant Professor _____ Associate Professor _____ Professor

   _____Other

3. **How many total years of collegiate aviation teaching experience do you have?**

   _____ 1-3 years_____ 4-6 years_____ 7-10 years _____ 10+ years

4. **How many total years of collegiate aviation teaching experience do you have?**

   _____ 1-3 years_____ 4-6 years_____ 7-10 years _____ 10+ years

5. **How many years of collegiate aviation teaching experience do you have at your current institution?**

   _____ 1-3 years_____ 4-6 years_____ 7-10 years _____ 10+ years

6. **What is your current tenure status?**

   _____ Tenured _____ Tenure track non-tenured _____ Not tenure track

7. Which primary area of instruction do you currently teach?

_____ Flight education _____ Aviation Management _____ Aviation Maintenance

_____ Aviation Electronics _____ Aviation Studies _____ Aviation Safety

_____ Air Traffic Control _____ Unmanned Aircraft _____ Graduate Studies

_____ Other

APPENDIX B

HUMAN USE EXEMPTION

LOUISIANA TECH
UNIVERSITY

OFFICE OF SPONSORED PROJECTS

EXEMPTION MEMORANDUM

TO: Kary Randall and Dr. Bryan McCoy

FROM: Dr. Richard Kordal, Director of Intellectual Properties
rkordal@latech.edu

SUBJECT: HUMAN USE COMMITTEE REVIEW

DATE: October 23, 2020

TITLE: "Assessing Student Ratings Regardless of Gender Equality of Aviation Faculty in Higher Education "

NUMBER: HUC 21-024

According to the Code of Federal Regulations Title 45 Part 46, your research protocol is determined to be exempt from full review under the following exemption category(s): 45 CFR 46.104(d) (2) i.

"Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) if at least one of the following criteria is met:

(i) The information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects;

(ii) Any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, educational advancement, or reputation; or

(iii) The information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects, and an IRB conducts a limited IRB review to make the determination required by § 46.111(a)(7)."

Thank you for submitting your Human Use Proposal to Louisiana Tech's Institutional Review Board.

APPENDIX C

CHANGING EVALUATION TERMINOLOGY TO CONSTRUCTS

Changing Evaluation Terminology to Constructs

Construct	Criteria	Examples	Literature Support
Expressiveness (RQ1)	Expressiveness refers to teaching behaviors related to faculty motivation in the classroom (Bennett, 1982; Stewart & Barraclough, 1992). Instructors who are dynamic and energetic by including hand gestures, smiling, and vocal inflections are considered expressive (Schonwetter et al., 1995) and influence student enthusiasm and interest (Holec & Marynowski, 2020; Murray, 1991).	-The instructor presented the course material in a manner that made it interesting. -The instructor created an environment that made helped students learn. -Shows interest in the subject matter. -The instructor showed enthusiasm for the subject matter. -The instructor was engaged while teaching the course.	Murray (1991) Holec & Marynowski (2020) Holec & Marynowski (2020) Murray (1991) Stewart & Barraclough (1992) Schonwetter et al. (1995) Holec & Marynowski (2020)
Immediacy (RQ2)	Immediacy refers to verbal and non-verbal behaviors linked with approachability and warmth (LeFebvre & Allen, 2014). A faculty member who regularly interacts with students and is available and caring has a high degree of perceived immediacy (Mehrabian, 1981; Stewart & Barraclough, 1992).	-The instructor was available to students outside of class. -The instructor seemed to care about our learning -The instructor was willing to answer questions during or outside of class. -The instructor was available during office hours. -The instructor interacted effectively with the students -The instructor was available for help outside of class.	Stewart & Barraclough (1992); Mehrabian (1981) LeFebvre & Allen (2014) Bennet (1982) Stewart & Barraclough (1992); Mehrabian (1981) Benntee (1982) Stewart & Barraclough (1992); Mehrabian (1981)

Construct	Criteria	Examples	Literature Support
Professionalism (RQ 3)	Professionalism refers to faculty behaviors that explain course material using structured lectures (Bennett, 1982; Feldman, 1993; Basow & Silberg, 1987; Marsh, 1987). Perceived knowledge of the subject area and competence is further evidence of faculty professionalism (Feldman, 1993).	- The instructor's course material was well-organized.	Bennett (1982); Feldman (1993) Feldman (1993); Marsh (1987)
		- The instructor communicated the subject matter clearly.	Bennett (1982); Basow & Silberg (1987)
		- The course material was delivered in a clear and organized manner.	Feldman (1993)
		- The instructor was knowledgeable about the subject.	Feldman (1993)
		-The instructor explains difficult material.	Bennett (1982)
		-The instructor was prepared for class.	Schonwetter (1995); Bennett (1982)
		-Instructor organizes and plans the course effectively.	
Openness (RQ4)	Openness refers to a faculty member's in-class adaptability and receptiveness to new ideas and questions in a classroom environment (Rubin, 1980). An instructor who exhibits openness encourages student participation and is receptive to new ideas (Feldman, 1993; Marsh, 1987; Rubin, 1980).	-The instructor encouraged student participation,	Marsh (1987)
		-The instructor provided opportunities for student participation.	Marsh (1987)
		-Instructor promoted active student participation and encouraged students to ask questions.	Rubin (1980); Marsh (1987); Feldman (1993)
		-The instructor encouraged an atmosphere in which I felt comfortable participating.	Rubin (1980)
		-I was encouraged to interact with the instructor regarding course content.	Bennett (1982)