Spring 2-2021

The Influence of Implicit Bias on Student Evaluations of Teaching at a Missouri Community College

Kimberly Franklin Berry
Lindenwood University

The Influence of Implicit Bias on Student Evaluations of Teaching at a Missouri Community College

by

Kimberly Franklin Berry

February 18, 2021

A Dissertation submitted to the Education Faculty of Lindenwood University

in partial fulfillment of the requirements for the degree of

Doctor of Education

School of Education

The Influence of Implicit Bias on Student Evaluations of Teaching

at a Missouri Community College

by

Kimberly Franklin Berry

This Dissertation has been approved as partial fulfillment

of the requirements for the degree of

Doctor of Education

Lindenwood University, School of Education

Sherry R. DeVore	February 18, 2021
Dr. Sherry DeVore, Committee Chair	Date
Kathy J. Grover	February 18, 2021
Dr. Kathy Grover, Committee Member	Date
Linda S. Caldwell	February 18, 2021
Dr. Linda Caldwell, Committee Member	Date

Declaration of Originality

I do hereby declare and attest to the fact that this is an original study based solely upon my own scholarly work at Lindenwood University and that I have not submitted it for any other college or university course or degree.

Full Legal Name: Kimberly Franklin Berry

Signature Kimberly Franklin Berry Date February 18, 2021

Acknowledgements

There are so many people to thank and acknowledge for helping and supporting me during this dissertation process. First, I would like to thank Dr. Sherry DeVore, my dissertation advisor. She guided me through the entire research and writing process. I will forever remember Dr. DeVore as encouraging, understanding, and patient. I am also grateful to the other members of my committee, Dr. Kathy Grover and Dr. Linda Caldwell, for their expertise and support.

Special appreciation is given to Dr. Vivian Elder for her compassion, words of wisdom, excellent writing, and proofing skills. Plus, I would like to thank my fellow colleagues, the participants, and all my current and past students for their encouragement and contributions.

Last but not least, my sincerest gratitude goes to my village: to the members of Pitts Chapel UMC and to my pastor, H. Russell Ewell, II; without their well-wishes and prayers, completing this study would not have been possible. To my husband, Richard, for his shoulders to lean on and his tough love during the difficult times. To my daughter, Gabrielle, for being my biggest cheerleader and not allowing me to quit or to doubt my ability to achieve success. To my mother, the late Helen Franklin, for believing in me and instilling the importance of education in me at an early age.

Abstract

Student Evaluations of Teaching (SET) are the primary instrument used to measure teaching effectiveness by colleges and universities nationwide (Wallace, Lewis, & Allen, 2019). Many colleges and universities use the SET for personnel decisions regarding tenure, promotion, and termination. Yet, the possibility of implicit bias impacting the SET and resulting in discriminatory practices against marginalized faculty continues to be a concern for higher education (Mitchell & Martin, 2018). The effectiveness of the SET continues to be debated by faculty and administrators due to concerns about reliability, validity, and bias (Bonitz, 2011). This quantitative study was conducted to examine the influence of implicit bias on the SET. The goal of this study was to determine if students were more influenced by instructor characteristics than teaching effectiveness when completing the SET and to determine how faculty perceived student responses on the SET. During the fall 2019 semester, faculty and students in the Communication and World Languages department at a Missouri community college were asked to complete a survey. The results from the student survey indicated the students were influenced by instructor characteristics, and they preferred native English-speaking instructors. The results from the faculty survey indicated the faculty believed students were influenced more by instructor characteristics than by teaching effectiveness. The findings in this study may serve as a reminder that the SET is influenced by implicit bias; therefore, marginalized groups may be negatively affected by SET results.

Table of Contents

Abstract……...iii

List of Tables……...……..…...viii

Chapter One: Introduction ……..1

Background of the Study ……....2

Theoretical Framework ……...4

Statement of the Problem ……5

Purpose of the Study ……..….8

Research Questions ……....9

Significance of the Study …….……...……9

Definitions of Key Terms …….……..…..…10

Delimitations, Limitations, and Assumptions …….……..11

Summary ……...13

Chapter Two: Review of Literature ……..…14

History of Student Evaluations…….…. ……...15

Theoretical Framework……...……..…….20

Healthcare…….…….…....21

Criminal Justice System……..……...22

Employment……...……23

Education……...24

Student Evaluations of Teaching……..…….25

Strengths of the SET…….……....26

Concerns about the SET……...…….….28

Online Surveys……..……...…….….31

Low Response Rate…….……...……32

Validity and Reliability……..……...…….…33

Leniency Bias……..……...……..…….….34

Student Characteristics……..……...…….….36

Marginalization……..……...……...…….…….37

Race and Ethnicity……..……...……..…….…….38

Gender……..……...…….…….…….40

Age and Physical Attractiveness……...……...…….…….44

English Proficiency ……..……...…….…….45

Summary ……...46

Chapter Three: Methodology ……48

Problem and Purpose Overview ……48

Research Questions ……...48

Research Design ……49

Population and Sample……..…49

Instrumentation …….....50

Reliability…….….50

Validity……..51

Data Collection…….….51

Data Analysis……...…….….52

Ethical Considerations…….……..…52

Summary……....53

Chapter Four: Analysis of Data…….54

Demographic Data…..……...55

Data Analysis……..…55

Student Demographic Statements……..……55

Faculty Demographic Statements……..……74

Summary……...…….90

Chapter Five: Summary and Conclusions ……...…….....91

Findings…….92

Research Question One……...……..92

Research Question Two……...…...……..93

Research Question Three……...…….……...95

Conclusions……..…….97

Research Question One……...……..…….97

Research Question Two……..…….……..99

Research Question Three…….…….……...101

Implications for Practice……..……..……..102

Recommendations for Future Research……..…….103

Summary…….…….104

References……106

Appendix A …….……122

Appendix B …….……126

Appendix C……..130

Appendix D……..131

Appendix E……..132

Appendix F……..133

Appendix G……..134

Appendix H.....135

Vita……...136

List of Tables

Table 1. *Student Respondent Demographics*……...……..……..56

Table 2. *Summary of Student Statement Five Responses by Gender*……..61

Table 3. *Summary of Student Statement Six Responses by Gender* ……....…….62

Table 4. *Summary of Student Statement Nine Responses by Race*…….…..……..66

Table 5. *Student Survey Statement 11 Responses*……....……..…….68

Table 6. *Student Survey Statement 14 Responses* ……...……..…….…….71

Table 7. *Student Survey Statement 16 Responses* ……..…..……..……..73

Table 8. *Faculty Respondent Demographics*……...……..……..74

Table 9. *Summary of Faculty Statement Five Responses by Gender*..……...…..79

Table 10. *Summary of Faculty Statement Six Responses by Gender* ……..……...….80

Table 11. *Summary of Faculty Statement Nine Responses by Race* …….……..83

Table 12. *Faculty Survey Statement 11 Responses* ……...……..……..…….85

Table 13. *Faculty Survey Statement 14 Responses*……...……..…….87

Table 14. *Faculty Survey Statement 16 Responses* ……...……...……..…….89

Chapter One: Introduction

Nearly 90% of North American post-secondary institutions use some form of the student evaluations of teaching (SET) instrument (Murray, 2005, p. 2; Wallace et al., 2019, p. 1). The term *SET* is used interchangeably with student evaluation of teaching effectiveness and student-instructor ratings and is the primary instrument to quantify the quality of an instructor's teaching and to obtain student feedback about a course (Perry, Wallace, Moore, & Perry-Burney, 2014; Sauer, 2012). The use of the SET is controversial due to the questionability of reliability and validity (Bonitz, 2011). This instrument is also troublesome to faculty and scholars because data are used for personnel decisions such as hiring, promotion, tenure, and termination (Basow, Codos, & Martin, 2013).

One common concern of faculty is the susceptibility of SET scores to bias and manipulation (Bonitz, 2011; Ray, 2018). Mitchell and Martin (2018) concluded the SET is systematically biased and can lead to discrimination in personnel decisions. As U.S. colleges and universities continue to change demographically and the recruitment of diverse faculty increases, administrators must be aware of potential problems with the SET (Smith & Hawkins, 2011).

Following a study conducted by Reid (2010), results indicated minority faculty were evaluated significantly lower than their white peers, which can be problematic when institutions demand excellent teaching for promotion and tenure. Mitchell and Martin (2018) found female professors are evaluated differently than male professors. Students refer to female professors as "teacher" and frequently comment on their appearance and personality, not their teaching (Mitchell & Martin, 2018).

Uttl, White, and Gonzalez (2017) determined the SET measures student happiness, which depends on many factors, including each student's satisfaction with the instructor's accent and looks. McPherson and Jewell (2007) suggested students rate younger instructors higher than older instructors, which may indicate a connection between perceived attractiveness and age. In other words, students may judge attractiveness as being a characteristic of youth (McPherson & Jewell, 2007). Even though experience was noted as a desired quality in faculty, bias regarding physical appearance was a factor in overall ratings (McPherson & Jewell, 2007).

This study was designed to measure the influence of implicit bias on the outcome of the SET. The variables examined in this study include ethnicity, race, gender, age, attractiveness, and English proficiency. The sections within this chapter contain historical information, the theoretical framework, a statement of the problem, the purpose and research questions, the significance of the study, definitions of key terms, and the delimitations, limitations, and assumptions of the study.

Background of the Study

The SET instrument has been used by colleges and universities to measure teaching effectiveness since the early 1920s (Degheri, 2017; Wachtel, 1998). The first SET instrument originated at Purdue University in 1927, and by the 1940s was used by many colleges and universities across the United States (Wachtel, 1998). Calkins and Micari (2010) reported that by the end of the 1950s, student power increased, which resulted in students no longer being "just a number" (p. 10).

By the end of the 1970s, student groups often had control of distributing and collecting the SET and sharing their thoughts with the campus community about the outcomes (Spooren & Christiaens, 2017). This shift created anxiety for non-tenured

faculty and led to faculty rallying and questioning the validity of the SET (Calkins & Micari, 2010). In 1974, the American Association of University Professors acknowledged student opinions about quality teaching were valid and important for learning (Calkins & Micari, 2010).

The initial intention of the SET was to help instructors become more effective in the classroom, but over time the SET became a tool to support tenure, promotion, transfers, termination, raises, and awards (Campbell, 2005). The SET continues to be used by institutions to demonstrate accountability of instructional quality to various stakeholders (Spooren & Christiaens, 2017). Two camps have formed, and one group defends the SET and asserts evaluations are true reflections of instructor effectiveness (Stroebe, 2016). The other group attacks the SET and asserts students do not have the experience or knowledge to evaluate teaching effectiveness (Carlozzi, 2017). McClain, Gulbis, and Hays (2018) proclaimed students might not be honest when completing the SET due to factors such as the timing of when the survey is administered and may feel pressured to complete assignments at the end of the semester, so they rush through the SET. Students may also respond dishonestly if they are uncertain their identities are protected (McClain et al., 2018).

In the 1980s, the concern of bias in student evaluations became a focus because more female professors entered college classrooms (Mitchell & Martin, 2018). Wallace et al. (2019) indicated male faculty were rated significantly higher than females for organizational skills, professionalism, and competence. Female faculty are also expected to be more nurturing and accessible than male faculty members (Mitchell & Martin, 2018).

Another area of concern is the impact of racial and ethnic bias on the SET. Research regarding ethnic and racial bias did not appear until the early 2000s, and the number of studies is minimal with mixed results (Calkins & Micari, 2010). Factors such as personality, age, accent, and perceived attractiveness are also new characteristics being researched (Wallace et al., 2019). Popular online publications of faculty ratings such as RateMyProfessors.com allow students to indicate if an instructor is "hot" or "not hot" in appearance (Coladarci & Kornfield, 2007, p. 2). Rosen (2017) found positive correlations between instructor quality and physical attractiveness. Braga, Paccagnella, and Pellizzari (2014) concluded the SET actually reflects customer satisfaction or likability, not the effectiveness of teaching. The debate over the reliability and validity of the SET as a measure of teaching effectiveness continues despite hundreds of studies about the issue (Stroebe, 2016).

Theoretical Framework

The psycho-sociological theory which provided the most appropriate framework for this study was Greenwald and Banaji's (1995) construct of implicit bias, also known as implicit social cognition. Staats, Capatosto, Tenney, and Mamo (2017) stated implicit bias includes "the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner. Activated involuntarily, without awareness or intentional control. Can be either positive or negative. Everyone is susceptible" (p. 10). Staats (2016) suggested implicit bias is not part of conscious awareness; however, the bias is pervasive and challenges individuals without explicit bias.

Jackson (2016) added, "Implicit associations can result in discrimination, even when people see themselves as egalitarian, and have no explicit intention to discriminate. People are often unaware that their unconscious associations can influence their

behavior" (p. 6). Staats (2016) further explained implicit bias "can challenge even the most well-intentioned and egalitarian-minded individuals, resulting in actions and outcomes that do not necessarily align with explicit intentions" (p. 29). Using Greenwald and Banaji's (1995) theory on implicit bias provided a means of examination and analysis of the unconscious bias students possess, which in effect, may increase the rate of negative responses regarding marginalized faculty.

Statement of the Problem

In light of the role the SET plays in tenure, promotion, faculty reputation, and teaching assignments, it is important to examine the perceptions of both faculty and students. Colleges and universities are held accountable for student success, which drives administrators to seek methods of collecting data to share with constituents as documentation of student performance (Hornstein, 2017). In this effort, students are surveyed regarding satisfaction with their learning experience, as discussed by Uttl et al. (2017):

> Typically, SET are conducted within the last few weeks of courses, before the final grades are assigned. Students are presented with rating forms that ask them to rate their perceptions of instructors and courses, often on a 5-point Likert scale ranging from Strongly Disagree to Strongly Agree. (p. 22)

Historically, the return rate of the SET is low, and the accuracy is questionable (Stark & Freishatat, 2014); therefore, more research regarding bias in the SET is critical due to potential consequences of negative ratings on faculty from marginalized groups. Mitchell and Martin (2018) concluded the SET is systematically biased and can lead to discrimination in personnel decisions. As U.S. colleges and universities continue to

change demographically and the recruitment of diverse faculty increases, administrators must be aware of potential problems with the SET (Smith & Hawkins, 2011).

A major concern regarding the reliability of the SET is leniency bias, which leads to instructors lowering expectations in the course to obtain positive evaluations (Gump, 2007). Hornstein (2017) suggested faculty feel pressure not to push students academically because of the potential consequences of negative evaluations. Coladarci and Kornfield (2007) noted instructors rated highly on easiness were also highly rated for overall quality in comparison to more difficult instructors. Several researchers have indicated students rate instructors higher when the grade expected in the course is good and punish instructors with low ratings when the grade expected in the course is bad (McClain et al., 2018; Wallace et al., 2019). Stroebe (2016) found teaching for positive evaluations may be in conflict with teaching effectiveness.

Calkins and Micari (2010) suggested the SET measures student satisfaction, not teaching effectiveness, which results in the continued debate regarding the validity of the SET to determine merit increases, tenure, and promotion. Furthermore, there is no consensus defining teaching effectiveness (Sauer, 2012). Hornstein (2017) believed the use of the SET would continue despite the possible problem of low return rates and biases that influence responses and subsequent interpretations of those responses. According to Hornstein (2017), administrators like to use the SET because of the ease and low cost.

In a study by Thielschi, Brinkmoller, and Forthmann (2018), completion rates were strongly correlated to student identification with the university and the course. Thielschi et al. (2018) suggested students may feel obligated to complete surveys if they are dedicated to the subject matter. In addition, Thielschi et al. (2018) noticed students

also complete the SET when there is a positive evaluation climate; therefore, institutions must promote and educate students about the relevance of the SET. Low response rates are a result of apathy, perceived lack of anonymity, and lack of importance (Berk, 2012). Hornstein (2017) stated:

> None of these reasons tend to be considered when particular university tenure and promotion committees interpret the scores. Instead, inevitably, the onus is on the faculty member being evaluated to justify "low scores" – a difficult and in many ways unjustified task since he/she does not have the relevant information on which to base an explanation, and there are significant questions as to the reliability of the instruments used to collect student evaluation information. (p. 4)

Feistauer and Richter (2018) determined the SET is also affected by how likable students find a teacher and not by teaching quality; student judgment regarding the likability of an instructor may be influenced by instructor reputation. Furthermore, Feistauer and Richter (2018) agreed instructor likability has a stronger impact on the SET than an interest in the subject matter, and "judgments of likability apparently change little after the first impression of a teacher has been formed" (p. 176).

Historically, college professors have primarily included white males, which creates a perception that white men are the most effective instructors (Bavishi, Madera, & Hebl, 2010). McPherson, Jewell, and Kim (2009) found SET scores are higher for white male instructors than for female and non-white instructors. Perry et al. (2014) determined students believe women and people of color are less intelligent than their white male counterparts.

According to Bavishi et al. (2010), gender and racial stereotypes may create different expectations for different individuals. Women are expected to be more

nurturing and caring than men, so women who did not possess those traits are rated more critically on the SET (Basow et al., 2013). Mitchell and Martin (2018) stated:

> Women have long claimed that their male counterparts are perceived as more competent and qualified. With mounting empirical evidence that this is true, perhaps it is time that universities use a method other than student evaluations to make these critical personnel decisions. (p. 652)

Recent court cases involving the University of Kansas and Pomona College are two examples in which professors filed discrimination cases due to employment decisions made by administrators based on the outcome of the SET (Schmidt, 2015, 2017). Smith and Hawkins (2011) concluded that non-white faculty consistently received lower ratings than faculty who indicated *Other* or *White* as their identity.

Purpose of the Study

The purpose of this study was to explore how instructor characteristics such as ethnicity, race, age, gender, accent, and personality influence the outcome of the SET at a Missouri community college. The study also served as a guide to investigate whether implicit bias influences how students respond to the SET. Recent researchers have indicated the SET often confirms the worse stereotypes about women faculty and faculty from marginalized groups (Basow et al., 2013; Mitchell & Martin, 2018; Ray, 2018). With the potential for administrators to misjudge faculty based on the outcomes of the SET, it is important to understand how instructor characteristics rather than the quality of instruction may be evaluated by students (Bavishi et al., 2010; Dev & Qayyum, 2017). Finally, the study served as a guide to investigate how faculty perceive the SET.

Research questions. The following research questions guided the study:

1. What percentage of students utilize a different criterion to evaluate faculty from marginalized groups than faculty from non-marginalized groups?
2. Based on the opinions of students and faculty, to what extent does implicit bias influence the outcome of student evaluations of teaching?
3. Based on the opinions of students and faculty, to what extent do instructor characteristics predict the outcome of student evaluations of teaching?

Significance of the Study

The literature examining the implications of bias in the SET instrument continues to grow; however, studies including race, ethnicity, and English proficiency are needed due to the continued use of the SET by administrators for personnel decisions (Mitchell & Martin, 2018). Also, despite the amount of research about the SET, the impact of bias is still inconclusive (Mitchell & Martin, 2018). There is a gap in the body of literature regarding the extent to which professor characteristics such as ethnicity, race, age, gender, and personality influence the outcomes of the SET (Mitchell & Martin, 2018).

Mitchell and Martin (2018) recommended further research to examine the effects of race, ethnicity, gender, and English proficiency on SET results, because the SET impacts employment decisions. Bavishi et al. (2010) asserted, "There is little research that addresses whether students hold stereotypes of professors based on ethnicity and gender and if these stereotypes lead to different judgments of professors" (p. 245). The research focused on "student judgments of professors is justified because students, either directly or indirectly, are an influence in a professor's career" (Bavishi et al., 2010, p. 246). Department heads, deans, and tenure committees need to be aware of the biased

nature of the SET to help reduce the harm inflicted due to reliance on such a biased measure (Ray, 2018).

Definitions of Key Terms

For the purposes of this study, the following terms are defined:

Ethnicity. Ethnicity differentiates Hispanic or Latino and non-Hispanic but allows individuals to select a race (U.S. Census Bureau, 2017).

Faculty. In postsecondary education, faculty is a term used to identify academic professionals, including professors, instructors, and lecturers employed full-time, part-time, or adjunct (National Center for Education Statistics [NCES], 2018; Stevens, Schneider, & Bederman-Miller, 2018). Faculty determine curriculum, control content, set student performance standards, and measure the outcomes of student performance (Echols, Neely, & Dusick, 2018; NCES, 2018).

Gender. For population purposes, the U.S. government defined gender as "a social construction whereby a society or culture assigns certain tendencies or behaviors the labels of masculine or feminine" (U.S. Census Bureau, 2018, para. 1).

Implicit bias. Implicit bias describes an unconscious or subconscious mental process in which one perceives individuals or groups of people who are different in a negative way (Staats, Capatosto, Wright, & Contractor, 2015). Implicit bias can result in discriminatory practices against individuals and groups (Staats et al., 2015).

Marginalization. Marginalization describes a systematic power imbalance directed toward specific groups based on ethnicity, gender, race, age, culture, and other differences (Causadias & Umaña-Taylor, 2018). Individuals from marginalized groups experience disadvantages and exclusion within society as a result of this social phenomenon (Causadias & Umaña-Taylor, 2018).

Perception. Perception is the process of selecting, organizing, and assigning meaning to the events, surroundings, and people one observes (Floyd, 2018). Perceptions are valuations and judgments of the world based on personal experiences and influenced by norms and values (Thijs & Verkuyten, 2016).

Race. Race describes how people identify according to social groups (U.S. Census Bureau, 2017). Persons can identify as belonging to multiple groups (U.S. Census Bureau, 2017).

Stereotyping. Stereotypes are a generalized set of beliefs about a group of people applied to individual members without consideration of possible differences (Floyd, 2018). Stereotyping is a way for people to organize information about groups different than their own (Floyd, 2018). These differences include "what members of a particular group look like, how they behave, or their abilities" (Türko, 2016, p. 54).

Student evaluation of teaching (SET). The SET is the instrument administered by a college or university to determine the effectiveness of an instructor's teaching (Uttl et al., 2017).

Delimitations, Limitations, and Assumptions

The scope of the study was bounded by the following delimitations:

Time frame. Data were collected during the fall 2019 semester.

Location of the study. The study took place at a Missouri community college.

Sample. Participants included adjunct and full-time faculty members of the Communication and World Languages Department and students enrolled in courses within the department during the fall 2019 semester.

The following limitations were identified in this study:

Population and sample demographics. The population in this study was limited to currently employed Communication and World Languages faculty at one Missouri community college. Student participants were currently enrolled in Communication and World Languages courses within the one community college in a multi-campus system. The sample in this study was further limited to members from both populations who responded to the survey used to collect data in this study. An additional limitation was the lack of diversity within the faculty pool. The population included less than 1% of faculty who identified as non-white or spoke English as a second language (M. Tollet, personal communication, January 18, 2019).

Instrument. The use of a survey for data collection in this study was a limitation. The survey in this study was delivered electronically through Qualtrics, and participants voluntarily completed the survey without supervision. These factors can result in unreliable data due to participants misinterpreting questions or simply answering dishonestly (Fraenkel, Wallen, & Hyun, 2019).

Researcher bias. Bias is defined as an inability to be objective about an issue or situation (*Oxford Online Dictionary*, 2019). The potential for researcher bias exists because of the researcher's background with the topic and relationships with the participants. The researcher may have been the instructor of some student participants, so she did not discuss the survey with her students. The researcher was also the department chair of Communication and World Languages and therefore had a supervisory role over the faculty surveyed for this research project. To address this potential bias, the researcher did not discuss the survey with faculty members within her supervision.

Summary

With nearly 90% of U.S. colleges and universities using the SET as a standard measurement of instructor performance, it is important to understand the history and problems surrounding its use (Murray, 2005, p. 2; Wallace et al., 2019, p. 1). Although numerous studies have been conducted to determine the reliability and validity of the SET, few researchers have examined the influence of professor characteristics (Bavishi et al., 2010; Feistauer & Richter, 2018). Understanding the implications of implicit bias in the student evaluations of teaching for marginalized faculty is important because administrators use the outcomes of the SET to hire, promote, terminate, and grant tenure (Boring, Ottoboni, & Stark, 2016; Campbell, 2005). Marginalized faculty are at a higher risk of receiving negative evaluations based on characteristics they cannot change and that do not correlate with their teaching effectiveness (Ray, 2018; Smith & Hawkins, 2011).

Within this chapter, the introduction and background regarding the controversy about the SET were presented. Implicit bias was identified as the theoretical framework for this study. A statement of the problem, purpose, research questions, and significance of the study were presented. Finally, key terms were defined, and delimitations, limitations, and assumptions were explained.

Contained in Chapter Two is a review of relevant literature. An in-depth examination of the most prevalent literature available about the SET is presented. The main topics of discussion include the impact of perceptions, stereotyping, and implicit bias on marginalized faculty.

Chapter Two: Review of Literature

Student Evaluations of Teaching (SET) are used by colleges and universities for assessing the quality of instruction (Young, Joines, Standish, & Gallagher, 2018). Initially, the purpose of the SET was to improve teaching, but over the years, the SET has been used to determine tenure and promotion (Stroebe, 2016). The SET process typically requires students to complete a survey during the last few weeks of a course (Uttl et al., 2017). Uttl et al. (2017) stated students are asked to rate the overall course and instructor using a Likert scale ranging from strongly disagree to strongly agree regarding instructor characteristics such as friendliness, fairness, enthusiasm, and availability. According to Murray (2005), the SET is designed to measure student learning by examining the quality of instruction. Murray (2005) added the SET measures course characteristics observed by students, applicable to various courses and controllable by the instructor. Bonitz (2011) explained, "A typical SET report contains descriptive statistics (mean, standard deviation, range, modal response, etc.) for the items scored on a scale, as well as the qualitative feedback provided by the students" (p. 16).

Proponents believe the SET measures effective teaching and provides students a voice, while others view the SET as biased, unreliable, and a major contributor to grade inflation (Bonitz, 2011). Carlozzi (2017) concluded, "The field of SET research can be understood as roughly divided between two opposing camps: (1) those who defend SETs, so-called 'SET apologists' and those (2) who attack SETs, so-called 'SET deniers" (p. 359). McPherson and Jewell (2007) emphasized concerns about the reliability and validity of the SET require the academic community to determine factors influencing the SET survey outcomes. In contrast, Rowan, Newness, and Tetradis (2017) stated:

> Because the SETs have been found to be valuable in assessing teaching effectiveness and are an integral component of the overall evaluation of faculty performance, we conclude that SETs should be used during personnel decisions, along with other factors such as scholarship research, authorship, faculty service, mentoring, growth and self-evaluation. (p. 1366)

Research about the SET is vast and continues to grow, but the findings and interpretation of the research have been inconsistent regarding the validity, reliability, and the impact of faculty characteristics (Carlozzi, 2017).

This review includes publications from 1927 to 2020, covering a range of more than 90 years of SET research; therefore, an extensive examination of literature was conducted regarding the SET. A review of the literature about the history of the SET, the theoretical framework which guided the study, the strengths of the SET, the concerns about the SET, and bias regarding SET results are presented in this chapter.

History of Student Evaluations

The original SET was created in the 1920s by Remmers from Purdue University (Stalnaker & Remmers, 1928). Remmers and his colleagues developed the Purdue Rating Scale for Instructors to examine the correlation between grades and evaluation scores (Sauer, 2012). Stalnaker and Remmers (1928) stated, "The Purdue Rating Scale for Instructors was developed to measure in an objective way the student opinion of the ability of an instructor for his task. All the traits measured are ones which an instructor may with effort alter" (p. 602). Ten qualities believed to exemplify effective teaching were measured by the Purdue Rating Scale:

> (1) interest in subject, (2) sympathetic attitude towards students, (3) fairness in grading, (4) liberal and progressive attitude, (5) presentation of subject matter,

> (6) sense of proportion and humor, (7) self-reliance and confidence, (8) personal peculiarities, (9) personal appearance, (10) stimulating intellectual curiosity. (Stalnaker & Remmers, 1928, p. 603)

Degheri (2017) related the SET dated back to the early 20th century and focused on "characteristics of intellectualism, individuality and equality associated with the period following World War I" (p. 5). Campbell (2005) explained student evaluations have evolved from the initial development stage, the voluntary participation era of the 1960s, the efficacy of ratings during the 1970s, and present-day focus on research regarding the validity of the SET.

In the early years, the SET was rarely administered to students and was not part of a formal process for evaluating teaching (Campbell, 2005). By the 1940s, the SET was used on a limited basis by institutions across the nation; however, instructors were not impacted by outcomes of the SET (Calkins & Micari, 2010). During this time, most professors were granted autonomy in the classroom due to the shortage of qualified faculty and strong support from administrators and the American Association of American Professors (Calkins & Micari, 2010).

Campbell (2005) noted by the early 1960s, the use of the SET increased as colleges provided faculty with surveys to administer to students at their discretion. Furthermore, according to Campbell (2005), "Faculty members managed the process with very little administrative involvement. Teachers who chose to use the student evaluation forms generally used the results for personal advisement on teaching practices" (p. 23). In the mid-1960s, students were distributing and collecting the SET around college campuses throughout the United States (Campbell, 2005). Calkins and Micari (2010) added:

> Student groups would administer and collect student ratings at their respective colleges, annotate them with often caustic or snide remarks, and circulate the annotated ratings widely through the university community so that they passed easily through the hands of department chairs, deans and faculty colleagues. (p. 11)

By the late 1960s, student ratings had become a source of anxiety for faculty and a "bitter contest" between faculty and administrators (Calkins & Micari, 2010, p. 7). Increasing demands by the federal government influenced colleges and universities to improve teaching and provide accountability as the shift toward consumerism emerged (Sauer, 2012). The anti-war movement surrounding the Vietnam War, along with the social unrest of the civil rights movement, contributed to the demand made by students to have a voice regarding their education (Degheri, 2017).

By the 1970s, "SETs were intended primarily for formative purposes, that is, to improve and shape the quality of teaching" (Hornstein, 2017, p. 2). Hornstein (2017) determined the SET evolved into a summative evaluation to measure overall performance and to determine promotion and tenure of faculty. The SET provided instructors with information about their strengths and weaknesses in the classroom; however, instructors had minimal department training or mentoring opportunities, so they typically implemented their own improvement plans (Otani, Kim, & Cho, 2012).

The 1970s was also a time when SET research emerged from a variety of disciplines, including English, engineering, and zoology (Kulik & Kulik, 1974). According to Kulik and Kulik (1974), psychologists laid the foundation for research about student ratings and determined the SET provided a "reliable, convenient, useful and probably valid method for evaluating teacher performance" (p. 51). However, the

findings of the psychologists were challenged by researchers from other disciplines, and the validity and reliability of the SET came into question (Kulik & Kulik, 1974).

The shift in viewpoint was evident following a study conducted by Rodin and Rodin (1972) wherein they examined a group of 300 students to determine the correlation between learning and the SET. Students were placed in 12 different sections of the same course with six different instructors; the content and structure of the classes were identical (Rodin & Rodin, 1972). The results of the study revealed the three lowest-rated instructors had students with the highest exam grades, and the instructors with the highest ratings had students with the lowest exam grades (Rodin & Rodin, 1972). Rodin and Rodin (1972) stated, "… perhaps students resent instructors who force them to work too hard and to learn more than they wish" (p. 1166).

In another study conducted by Natfulin, Ware, and Donnelly (1973), an actor introduced as "Dr. Fox" presented a non-substantive lecture in an engaging and charismatic way to an audience of mental health professionals. The audience members were instructed to evaluate "Dr. Fox," and the results indicated audience approved of the lecture content and the lecturer (Natfulin et al., 1973). Natfulin et al. (1973) concluded the ratings were impacted more by instructor likability and presentational style than by the content quality. Through the years, researchers have continued to examine the concept of likability as a potential threat to the validity of the SET (Natfulin et al., 1973).

In a study conducted by Feistauer and Richter (2018), students completed a questionnaire on the first day of class about their interest in the subject, prior knowledge of their instructor, and likability of their instructor. Several weeks later, those same students completed the SET, and results indicated a correlation between instructor likability and instructor ratings (Feistauer & Richter, 2018). Feistauer and Richter (2018)

concluded, “SETs are affected by strong biasing effects of how likable students find a teacher and by weak biasing effects of how strongly they are interested in the course subject” (p. 177). Feistauer and Richter (2018) also asserted the students’ first impression of the instructor did not change during the course; therefore, instructor behavior had minimal impact on SET ratings.

Studies continued to emerge during the 1970s and 1980s as researchers attempted to determine the effectiveness of the SET for measuring teaching performance (Wallace et al., 2019). Wallace et al. (2019) determined, “Views are mixed regarding the validity of SETs, with evidence for and against their use as measures of instructor teaching performance” (p. 1). By the mid-1980s, the SET became a popular tool for measuring teaching effectiveness (Stroebe, 2016). Huston (2006) noted the number of female faculty members also increased during this time, as the number of doctorates awarded to women increased by nearly 50% between 1975 and 2001 (p. 594). Female faculty members expressed frustration with male colleagues, administrators, and students due to differing expectations for women (Boring et al., 2016). These concerns and questions led to research about potential biasing factors regarding gender and the SET (Basow et al., 2013).

By the early 1990s, the SET had become the primary instrument for obtaining student feedback about courses and instructor effectiveness (Perry et al., 2014). Colleges and universities began the transition to the online SET rather than the in-class-administered SET (Young et al., 2018). Young et al. (2018) explained the online SET allows all students the opportunity to complete the surveys since absenteeism on the day of the SET is no longer a concern, plus online surveys eliminate the possibility of faculty recognizing student handwriting for written responses.

In 2000, approximately 2% of U.S. institutions had transitioned from paper and pencil evaluations to online, and by 2005 nearly 33% had made the transition (Anderson, Brown, & Spaeth, 2006, p. 1). McClain et al. (2018) determined online surveys had risen in popularity due to environmental consciousness regarding paper waste, and the reduction in campus budgets demanded universities find ways to save money. More recently, research about the SET has focused on expanding earlier studies of potential biasing factors regarding students, faculty, and the course (Sauer, 2012).

Theoretical Framework

The theoretical framework most suitable for this study was implicit bias, also known as implicit social cognition (Greenwald & Banaji, 1995). Implicit bias is an unconscious or subconscious mental process in which one perceives individuals or groups of people who are different in a negative way (Staats et al., 2015). Greenwald and Banaji (1995) developed the theory of implicit bias during the 1990s when they created the implicit bias test in which they displayed pictures of people from various racial groups and examined the reactions or preferences displayed by those taking the test. Banaji and Greenwald (2013) suggested implicit bias influences all people regardless of explicit attitudes. Payne, Niemi, and Doris (2018) added, "This tendency for stereotype-confirming thoughts to pass spontaneously through our minds is what psychologists call implicit bias. It sets people up to overgeneralize, sometimes leading to discrimination even when people feel they are being fair" (p. 2).

People develop implicit bias as young children through exposure to images and learned behaviors from their communities (Tyner, 2019). Tyner (2019) further stated, "The ideas and images over time become a part of our perspectives and influence us even when we do not realize it. These instances are manifested in our verbal/nonverbal

communication, body language, and everyday life" (para. 10). For example, Price and Payton (2017) concluded from their study some police officers may have a propensity to discriminate against African-American males and expect criminal behavior. Price and Payton (2017) believed implicit and explicit bias contributes to the use of excessive force by police and disproportionate incarceration of African-American males.

According to Rynders (2019), decision-makers with a high level of discretion and a low likelihood of being reviewed by others are likely to be influenced by implicit bias and to act in a discriminatory manner. FitzGerald, Martin, Berner, and Hurst (2019) stated, "There has been much recent interest in studying the effects of implicit bias have on behavior; particularly when that may lead to discrimination in significant areas of life, such as health care, law enforcement, employment, criminal justice and education" (p. 2). Areas in society affected by implicit bias are discussed in the following sections.

Healthcare. In the area of healthcare, implicit bias can unintentionally impact the diagnosis and treatment of patients from marginalized groups (FitzGerald et al., 2019). Staats, Capatosto, Wright, and Contractor (2013) reported the quality and type of care provided varies by the patient's race. For example, Sabin and Greenwald (2012) found lower dosages of pain medication were prescribed to patients from marginalized groups than to whites. In addition, Merino, Adams, and Hall (2018) found diagnosis and treatment of mental health conditions are impacted by implicit stereotypes. Healthcare professionals are more likely not to complete assessments, misdiagnose, and over-diagnose psychotic disorders of marginalized groups (Merino et al., 2018).

FitzGerald and Hurst (2017) suggested already vulnerable populations, including immigrants, those with low income, the overweight, and minorities, are often subject to negative interactions with healthcare professionals. Communication between doctors and

patients also differs; researchers have discovered doctors dominate verbal communication with non-white patients, and non-white patients feel disrespected and ignored (Staats et al. 2013). Dehon et al. (2017) reviewed nine studies involving physicians and patient care to determine the relationship between clinical decision-making and racial bias. Results indicated most doctors implicitly prefer whites to blacks; however, in seven of the nine studies, bias did not impact clinical decisions (Dehon et al., 2017). Merino et al. (2018) concluded, "Implicit biases can negatively influence a provider's willingness to engage in patient-centered care, provide referrals to specialized treatment, or even adhere to evidence-based guidelines when serving diverse populations" (p. 723).

Criminal justice system. Kovera (2019) explained racial disparities in law enforcement continue to impact policing and the prison population. Kovera (2019) determined blacks are more likely to be stopped for traffic violations, be subjected to searches, be arrested, and experience excessive force. Explicit bias and stereotyping are a factor; however, implicit bias screening of police, judges, and attorneys suggests there is an unconscious awareness and lack of intent to treat groups differently (Kovera, 2019). Rynders (2019) found:

> … research on implicit bias in public defender decision-making shows that implicit bias can: (1) affect evaluation of ambiguous evidence, (2) influence how attorneys interpret a client's ambiguous behaviors and facials expressions, (3) negatively influence attorneys' behaviors and (4) cause attorneys to treat stereotyped individuals in stereotype-consistent ways. (p. 464)

Spencer, Charbonneau, and Glaser (2016) went on to say biased policing is most likely a result of implicit bias, because humans tend to unconsciously rely on stereotypes to judge ambiguous situations.

According to Spencer et al. (2016), police officers are normal human beings and are therefore subject to the influence of societal stereotypes. Spencer et al. (2016) concluded:

> Because they are often under conditions of uncertainty, high discretion, and stress and threat, the pervasive stereotypes linking Blacks and Latinos with violence, crime, and even specifically weapons are likely to cause them to make misattributions in seeking to disambiguate the intentions and behaviors of citizens. This can lead to racially disparate rates of stops, searches and use of force. (p. 59)

Whitfield (2019) advised growing up with varying cultural experiences results in a racial divide in which individuals, including police officers, rely on perceptions, not on facts. These perceptions may lead white officers to fear the natural behaviors of other cultures and therefore overact during interactions with people of color (Whitfield, 2019).

Employment. Since the Civil Rights Act of 1964, workplace discrimination has evolved from deliberately discriminatory acts to subtler, nearly undetectable forms of behavior (Jones, 2017). Jones (2017) believed marginalized groups experience disparate impact (unintentional discrimination) created by implicit bias. For example, plaintiffs brought a disparate treatment and impact case against Wal-Mart, claiming their policy to allow local managers decision-making for promotions and pay increases resulted in a discriminatory outcome for women due to Wal-Mart's gender-biased culture (Oncidi, 2018).

Implicit bias in the workplace occurs in applicant screening, interviews, retention, and even in providing health care services (Khush, 2020). Implicit bias may begin within the job posting; studies have shown the language used in postings may reflect the

characteristics of men resulting in a reduced number of female applicants (Legault, 2019). Legault (2019) clarified words such as competitive, outspoken, strong, and confident attract men, and words such as supportive, cooperative, and collaborative attract female candidates. In the screening of applicants, implicit bias may play a role as screeners unconsciously avoid applicants with ethnic or foreign-sounding names (Khush, 2020). For example, in a study conducted at Harvard, a female applicant received numerous invitations for job interviews when using her spouse's Anglo-Saxon last name but only one invitation when using her foreign-sounding maiden name (Khush, 2020).

Education. The influence of implicit bias impacts education in multiple areas, including assessment, discipline, and disproportionate representation in special education classrooms (Rynders, 2019). Fiarman (2016) recounted asking a visiting educator to collect data about the difficulty level of questions asked of students during the class discussion; results revealed questions requiring more critical thinking were asked of white students, and lower-level questions were directed toward students of color. Fiarman (2016) added the institution used this information as an opportunity to understand how bias impacts interactions with students.

Annamma and Morrison (2018) argued dysfunctional education systems are maintained through implicit bias; therefore, marginalized students, specifically disabled students of color, are at risk. According to Annamma and Morrison (2018), disabled students of color are more likely to be suspended for disciplinary issues than their white or non-white peers without disabilities, indicating an influence of implicit bias for race and disability. Nance (2019) suggested schools must move toward a "more equitable and inclusive academic environment" (p. 102). In addition, Nance (2019) explained implicit

bias has resulted in the use of unequal surveillance and punishment of marginalized groups.

Student Evaluations of Teaching

Researchers have clearly shown student perceptions of faculty characteristics play a role in how students rate the effectiveness of teaching (Wallace et al., 2019). Understanding the impact of implicit bias on the SET is important for improving the interpretation of SET results, and relying only on the SET to determine the overall effectiveness of faculty can result in discriminatory practices and can damage faculty members' careers (Boring et al., 2016). Reinsch, Goltz, and Hietapelto (2020) explained, "If implicit bias involves any of the protected categories under the law and evaluations are used to make employment decision, then those employment decisions are based on some factors that are discriminatory and therefore illegal" (p. 116). Students should be educated on the importance and the purpose of the SET (Wallace et al., 2019). Wallace et al. (2019) stated:

> Also, there is a need for online content-analysis methods that can read and detect bias in written SEs. Doing so may also provide evidential proof to reveal how faculty members' demographic characteristics may affect comments received on these summative instruments. (p. 9)

Rowan et al. (2017) asserted progress in design, collection, and interpretation of the SET is needed to ensure validity and reliability.

Uttl et al. (2017) emphasized colleges must first determine if their institutional focus is student learning or student satisfaction. The SET studies indicate professors who earn the highest ratings do not have students with the highest grades; therefore, SET

results may be a better indicator of student satisfaction, not learning (Uttl et al., 2017). Uttl et al. (2017) suggested:

> Universities and colleges focused on student learning may need to give minimal or no weight to SET ratings. In contrast, universities and colleges focused on students' perception or satisfaction rather than learning may want to evaluate their faculty's teaching using primarily or exclusively SET ratings, emphasize to their faculty members the need to obtain as high SET ratings as possible (i.e.; preferably the perfect ratings), and systematically terminate those faculty members who do not meet the standards. (p. 40)

Reinsch et al. (2020) concluded administrators believe effective teaching means the faculty members earn an above-average score on the SET. The authors concluded, "this makes it difficult for certain groups of people—usually underrepresented group members—to achieve 'above average' ratings while making it easier for members of majority groups to do so" (Reinsch et al., p. 116). McClain et al. (2018) noted SET results are also impacted by untruthful student responses. Several researchers have indicated students admit not being truthful with both scale questions and written comments (Clayson & Haley, 2011; Reynolds, 1977; Sproule, 2000).

Strengths of the SET. The primary purpose of the SET is to quantify the quality of a course and instruction (Perry et al., 2014). Stalnaker and Remmers (1928) stated, "The ratings, of course, are anonymous. Their purpose is to give the teacher interested in self-improvement an opportunity to get an objective check on the student opinion of his ability as an instructor" (p. 603). The SET provides students with an opportunity to anonymously voice opinions about their learning experience to faculty and administrators (Linse, 2017). McKeachie (1990) determined the SET provides students the opportunity

to express what behaviors they observed from their instructors and how those behaviors impacted their learning experience. Despite the continued reliability debate, the SET continues to be the best tool for collecting data regarding student perceptions of learning experiences, according to Spooren and Christiaens (2017). Centra (2003) concluded:

> No method of evaluating college teaching has been researched more than student evaluations, with well over 2,000 referenced in the *ERIC* system. The preponderance of these study results has been positive, concluding that the evaluations are: (a) reliable and stable; (b) valid when compared with student learning and other indicators of effective technology; (c) multidimensional in terms of what they assess; (d) useful in improving teaching; and (e) only minimally affected by various course, teacher, or student characteristics that could bias results. (pp. 495-496)

Gump (2007) explained the SET provides data for researchers and instructional improvement recommendations for faculty.

Uttl et al. (2017) concluded from their review of literature that proponents believe the SET provides accountability to stakeholders; the SET is convenient, inexpensive, and provide students a voice. In addition, Uttl et al. (2017) also stated:

> The SET proponents assume that students observe instructors' behavior, assess how much they learned from the instructor, rate the instructor according to how much the instructor's contributed to their learning, and thus, high correlation between SET and measures of learning should follow. (pp. 22-23)

Rowan et al. (2017) added, "Students, as the recipients of instruction, are capable of evaluating teaching effectiveness based on their perceptions and experiences and thus can

provide meaningful formative and summative feedback to instructors and administration" (p. 1364).

Rowan et al. (2017) clarified students are consumers of a product; therefore, their opinions should be included in personnel decisions. Meaningful data from the SET can increase faculty awareness of student concerns and ultimately improve teaching while addressing student needs (Rowan et al., 2017). Linse (2017) agreed the appropriate use of the SET can enable faculty to achieve a greater understanding of student perceptions and help administrators obtain the mission of their institution. According to Linse (2017), "The majority of the legitimate research on student ratings indicates that they are a more reliable and valid representation of teaching quality than any other method of evaluating teaching, including peer observations, focus groups and external review materials" (p. 97).

Rowan et al. (2017) asserted quality student responses increase when administrators communicate to students their input is valued and taken seriously. The SET is a vital source of information when constructed properly, and it provides information for improving classroom instruction and personnel decisions (Coladarci & Kornfield, 2007). The inclusion of SET results as part of faculty evaluations encourages instructors to improve teaching, address student concerns, and adopt varied delivery styles (Kornell & Hausman, 2016).

Concerns about the SET. More than 90% of colleges and universities use the SET despite the growing concern regarding reliability and validity (Wallace et al., 2019, p. 1). Stroebe (2016) stated there is growing concern regarding how administrators use the data obtained for the SET to determine personnel decisions such as merit increases, awards, and leadership roles. Some faculty perceive the SET as threatening to their

careers and reputations due to a multitude of articles about bias and unreliable results (Linse, 2017). Linse (2017) explained academic news sources such as *Inside Higher Education* and *The Chronicle of Higher Education* often publish opinion pieces about the SET, which are not peer-reviewed articles. These stories incite fear and anxiety in faculty about the unfair use of student ratings by administrators (Linse, 2017).

Otani et al. (2012) noted faculty members are evaluated on variables not within their control, such as class size, whether the course is a requirement, prior interest in the topic, difficulty level of the curriculum, mode of delivery, and instructor characteristics. Hornstein (2017) asserted the SET hinders academic freedom because faculty may avoid controversy and slow the pace of coursework to reduce student resentment, which could result in retaliation on the SET. Instructors may feel pressured to alter pedagogy in an effort to provide more entertainment or lessen expectations to receive higher ratings (Rowan et al., 2017).

The SET has limitations because students cannot assess non-classroom tasks such as assignment quality, course design, or instructor knowledge (Murray, 2005). Hornstein (2017) added, "The validity of anonymous students' evaluations rests on the assumption that, by attending lectures, students observe the ability of the instructors, and that they report it truthfully" (p. 3). Hornstein (2017) also expressed concern about the interpretation of the measurement:

> … few administrators are trained to interpret SET data. It is not uncommon for administrators to examine the scores and assume that those below the mean are bad and those above it are good – never mind that the calculation of means in these situations are simply inappropriate and meaningless. (p. 3)

Stark and Freishatat (2014) found administrators compare instructor scores to the department average in an effort to determine the effectiveness of teaching, but those comparisons are meaningless because "the numbers are labels, not values" (p. 2). The presumption that the number three holds the same meaning to different students in different classes is not valid, because "SET scores are ordinal categorical variables" and averaging the scores is senseless (Stark & Freishatat, 2014, p. 2).

Many opponents of the SET have clarified SET results represent student satisfaction with the instructor and the course, not the level of learning or teaching effectiveness (Bonitz, 2011; Rowan et al., 2017). Stroebe (2016) introduced the concept of revenge and reciprocity, which simply states students with poor grades evaluate instructors harshly in retaliation. According to Stroebe (2016), "Because students are aware that SETs are used in the evaluation of teachers by department heads or deans, some particularly angry students might give poor ratings in the hope that it will have negative consequences for their instructor" (p. 804).

The concept of reciprocity is the opposite of revenge and results in students with good grades rewarding faculty with positive ratings (Stroebe, 2016). A meta-analysis of multi-section SET studies conducted by Uttl et al. (2017) yielded no correlation between learning and SET ratings. In some cases, instructors with high ratings had students with poor grades, which suggests surveys may reflect student satisfaction, not teaching effectiveness or learning (Uttl et al., 2017). Prasad, Ko, and Sanchez (2017) explained:

> "Consumerism" in higher education is one framework that may explain the lack of relationship between SET ratings and student learning. In this framework, students behave like consumers of a commodity. Studies have shown that

> teachers who are more "likable" receive disproportionally higher SET scores. (p. 1367)

Students decrease ratings for instructors perceived as boring and increase scores for instructors perceived as interesting (Prasad et al., 2017).

Online surveys. Online surveys may be popular but are not free of controversy (McClain et al., 2018). McClain et al. (2018) found students report not being honest when completing online evaluations because they are not convinced their identities are protected in the online format. Online surveys also tend to have a lower response rate because student participation is voluntary (Spooren, Brockx, & Mortelmans, 2013). Students have also indicated they do not complete surveys due to time constraints and because they see no personal benefit for completing surveys (Spooren & Christiaens, 2017). Young et al. (2018) discovered online SETs are completed at a lower rate than pencil-and-paper surveys and at a higher rate by dissatisfied students. Furthermore, Young et al. (2018) determined online surveys with a lower response rate could potentially create bias if the segment of students completing the surveys is not a representation of the entire class.

Another controversial online survey is Ratemyprofessors.com, a review website launched in 1999 (Rosen, 2017). RateMyProfessors.com allows students to comment on instructor clarity, helpfulness, and easiness (Coladarci & Kornfield, 2007). Rosen (2017) noted students also answer an optional question regarding the hotness or physical attractiveness of their instructors. According to Rosen (2017), "If a professor has a *hotness* score that is greater than 0, the professor is considered 'hot,' and an image of a chili pepper is displayed on the professor's Rate My Professors profile" (p. 3). Flaherty (2018) stated, "Female professors and their supervisors also report that open-ended

comments in end-of-term student evaluations too often disparage or otherwise focus on women's appearances" (p. 3). In conclusion, Rosen (2017) explained professors rated as *'hot'* score higher Ratemyprofessors ratings than those not considered 'hot,' and overall ratings are higher for easiness. Coladarci and Kornfield (2007) added, "Instructors deemed hot have somewhat higher ratings on both overall quality and easiness when compared to those who do not enjoy this distinction" (p. 3).

Low response rate. Low response rate is the lack of student responses on the SET (Lawrence, 2018). Some researchers have asserted the problem is growing due to the transition from paper-pencil surveys to optional online surveys (Young et al., 2018). Lawrence (2018) explained some administrators attribute low response rates to faculty, but there is no evidence to support this argument.

Young et al. (2018) suggested the low response rate is an issue of lack of student motivation to complete the SET, because students perceive their responses have no impact on administrators or faculty. According to Thielschi et al. (2018), SETs with a low response rate indicate only students interested in the course complete the surveys, so the opinions of other students are absent from the results. Furthermore, Thielschi et al. (2018) stated, "This causes more accented data patterns; in SET with lower participation, lecturers with positive evaluations receive even better scores and lecturers with bad evaluations are rated worse" (p. 189).

Young et al. (2018) conducted a study to determine if administering surveys during class time increases response rates. Participating faculty were advised not to offer incentives but to remind students of the date for the SET, encourage them to bring electronic devices on that day, and offer additional completion opportunities if they missed class (Young et al., 2018). According to Young et al. (2018):

> Our approach, using a quasi-experimental nonequivalent groups design to test the effect of a particular combination of tactics, lead to a 27.3 average percentage point increase in response rates, which is incredibly large given the starting average response rate was only 44.2%. (Young et al., 2018, p. 45)

Results indicated faculty reminders communicated to students the importance of the completion of the SET (Young et al., 2018).

Validity and reliability. Research results regarding the validity and reliability of the SET to measure teaching effectiveness have been inconclusive and inconsistent despite years of study on the topic (Braga et al., 2014; Spooren et al., 2013; Stroebe, 2016). Potential bias in the SET and a lack of student understanding of the importance of the SET contribute to concerns about its validity and reliability for measuring teaching effectiveness (Spooren et al., 2013). Uttl et al. (2017) stated:

> The opponents of SET as measures of teaching effectiveness argue that SET have no or only limited validity as a measure of instructor teaching effectiveness because both SET and measures of learning are influenced by teaching effectiveness irrelevant factors (TEIFs) such as academic discipline/field of study, student interest, student motivation, instructor sex, instructor accent, class level, class size, class meeting time, etc. (p. 23)

Dev and Qayyum (2017) concluded the SET is not reliable because students do not complete survey forms, thoroughly read the questions, or consider the process in a responsible manner. McClain et al. (2018) pointed out SET results are unreliable because students do not understand the purpose is to benefit themselves and future students.

Spooren and Christiaens (2017) conducted a study to explore the correlation between SET scores and student perceptions of the SET. Results indicated students

perceive value in the SET for accountability of teaching quality, but they doubt faculty members actually make changes in their teaching style based on student comments (Spooren &Christiaens, 2017). Spooren and Christiaens (2017) stated:

> We found a statistically significant relationship between SET scores and students' perceived value of SET practice. Students who value SET procedures tend to provide higher SET scores. This might confirm our hypothesis that strongly engaged students are more positive about all activities in their institution, such as teaching and SET practice. (p. 48)

Otani et al. (2012) found students often do not experience changes in future courses, which may perpetuate dishonest feedback and a low response rate on the SET. In addition, instructors may not have professional development opportunities to address weaknesses, or they may not know how to prioritize, so they rely on their interpretation of data to select areas needing improvement (Otani et al., 2012).

McClain et al. (2018) conducted a study to determine the honesty level of student responses on the SET and student understanding of the purpose of the SET. Results indicated students are more honest if they believe their responses directly impact the administrators' decision-making regarding retention and dismissal of faculty (McClain et al., 2018). Students are more honest when the SET is administered at the end of the semester rather than in the middle of the semester (McClain et al., 2018). McClain et al. (2018) explained students might be more honest at the end of the semester because they have experienced the entire course; therefore, their perspectives regarding the value of the course change, plus concerns about anonymity lessen once grades have been posted.

Leniency bias. The grade point average at colleges and universities has steadily increased since the 1980s, even though students spend less time on academics (Stroebe,

2016). Wachtel (1998) defined leniency bias as the practice of lowering grading standards in an effort to improve instructor ratings. Stroebe (2016) suggested students rate faculty based on grade expectations, which leads faculty to inflate grades and lower expectations to improve overall SET ratings. In the words of Stroebe (2016):

> According to the bias assumption, the work students are required to invest in a course and the grades they receive biases their evaluation of course and instructor. The less work students have to do and the better the grade they receive, the more positive their teaching evaluation. (p. 801)

Stroebe (2016) explained grade inflation is possible without bias, and only the perception of bias is needed. In other words, instructors may lessen requirements and grade more leniently because they believe students will reward them with positive evaluations (Stroebe, 2016).

In a study by McPherson et al. (2009), data from 24 consecutive semesters of Principles of Economics courses at the University of North Texas were analyzed to determine if there was a correlation between SET ratings and grade expectations. McPherson et al. (2009) stated," . . . instructors can 'buy' higher scores by increasing the grade expectations of their students; specifically, inflating students' expected grade by one letter grade would cause an instructor's evaluation score to rise by 0.2714 points" (p. 43). Lawrence (2018) added, "Professors are rewarded for being less demanding and more lenient graders both by receiving favorable SET ratings and by enjoying higher student enrollment in their courses" (para. 6).

Instructors with low ratings also inadvertently decrease their workload and a student's workload when they attempt to provide more class time for difficult projects to improve ratings and reduce negative comments; unfortunately, this practice can result in

deleting course content because there is no longer time to include everything (Lawrence, 2018). Gump (2007) examined 70 years of leniency hypothesis studies and discovered a lack of consensus regarding the existence of leniency bias and mixed results regarding the impact of leniency bias on SET outcomes. Gump (2007) added, "A new wave of research on SETs seems to be in order: research that looks critically yet holistically at past studies with respect to their methodologies, conclusions, and implications" (p. 66).

Student characteristics. Student characteristics including age, gender, interest in the course, and expected grade, influence responses on the SET (Wallace et al., 2019). For example, first-year students provide the lowest ratings on the SET, while older students rate older male instructors the highest on the SET (Degheri, 2017). Wachtel (1998) asserted the greater the student interest in a subject, the higher rating granted on the SET.

Feistauer and Richter (2018) examined the possible biasing effects of prior subject interest on the SET and found minimal impact. The authors determined, "In sum, the majority of previous studies found rather weak relationships of prior subject interest and SETs, which suggest that prior subject interest exerts a consistent but relatively harmless bias that only slightly compromises the validity of SETs" (Feistauer & Richter, 2018, p. 177). In an examination of undergraduate economics courses, McPherson et al. (2009) found an increase in SET ratings in sections with greater numbers of economics majors, which suggests an interest in the subject does influence SET ratings. McPherson et al. (2009) also noted sections with more female than male students granted higher SET scores. McClain et al. (2018) concluded further research is warranted to determine the relationship between student characteristics and the SET due to lack of research and mixed results reported by researchers over the past decade.

Marginalization

Bias in the SET is a growing concern regarding reliability and validity when assessing the teaching effectiveness of marginalized groups (Fan et al., 2019). Bias is defined as an inability to be objective about an issue or situation (*Oxford Online Dictionary*, 2019). According to Causadias and Umaña-Taylor (2018), marginalization described a systematic power imbalance directed toward specific groups based on ethnicity, gender, race, age, culture, and other differences. According to Bavishi et al. (2010), women and minorities experience marginalization in academia due to societal stereotypes and perceptions. For example, the idea of a professor may create a picture of a white male, not a female or ethnic minority (Bavishi et al., 2010). Stereotypes about African Americans might include not deserving their positions and only teaching due to affirmative action (Wallace et al., 2019). Other common stereotypes include extreme competence of Asians and sensitivity and kindness in women (Wallace et al., 2019). Lawrence (2018) suggested these instructor characteristics create substantive concerns because of student responses on the SET.

Early researchers Stalnaker and Remmers (1928) acknowledged the potential of biasing factors: "The halo effect in the teacher rating scale would mean that a student who likes a teacher for any reason whatsoever, therefore rates him high in all traits, even those in which he actually is deficient" (p. 606). McPherson et al. (2009) found that male, white, and younger instructors receive a significantly higher SET score than female, non-white, and older instructors. Fan et al. (2019) added faculty from marginalized groups, including non-English language backgrounds and women, are subject to varied biases. According to Reinsch et al. (2020), "The literature on performance appraisal clearly backs up our assessment that many items in teaching

evaluations are formed in a way that encourages or elicits, rather than discourages, the application of stereotypes to evaluating performance" (p. 123).

The SET rarely has statements about the behaviors of the instructor but may include opinions or perceptual items (Reinsch et al., 2020). For example, "The instructor displayed in-depth knowledge about the subject" requires the student to make a value judgment about the level of expertise an instructor has when the student has no expertise in the subject (Reinsch et al., 2020, p. 123). Ultimately, relying on the SET to determine faculty employment decisions may lead to discriminatory practices and litigation for colleges and universities (Reinsch et al., 2020). Linse (2017) summarized, "Inappropriate use of student ratings breeds mistrust, fosters inequities and inconsistencies, and ultimately demoralizes the faculty" (p. 103).

Race and Ethnicity

Quantitative research about the influence of race on the outcomes of the SET is limited and inconclusive (Huston, 2006). Smith and Hawkins (2011) stated, "The most noted and unexplored adverse situation are problems associated with teaching evaluations and diverse faculty, particularly Black faculty" (p. 149). A literature review by Smith and Hawkins (2011) revealed few studies involving exploration of the impact of race and ethnicity on the outcomes of the SET by comparing SET ratings of Asian, African American, and Hispanic faculty to white faculty (Anderson & Smith, 2005; Hamermesh & Parker, 2005; Smith, 2007).

Reid (2010) stated minimal empirical research has been conducted on race, even though the racial diversity of students and faculty continues to increase across college campuses. In addition, until recently, SET research about the impact of race has been limited to comparisons among faculty who are Hispanic, Asian, and White (Fan et al.,

2019). According to Basow et al. (2013), research in the area of race is more limited than research about gender due to the low number of non-white professors in comparison to white professors. Bavishi et al. (2010) explained barriers such as social isolation, slower rates of advancement, and a lack of academic mentors have contributed to the small percentage of faculty of color in colleges and universities. Basow et al. (2013) asserted research focused on the impact of race is difficult due to the inability to eliminate other variables including gender, attractiveness, personality, and teaching experience. As stated by Smith and Hawkins (2011), the increase of racially and culturally diverse students has led to the increase of more diverse faculty, and those faculty members contend students' ratings fail to reflect their teaching performance.

In a study conducted by Reid (2010), Ratemyprofessors.com data from the top 25 liberal arts institutions listed on the 2005 *U.S. News and World Report* rankings were examined for racial and gender bias (p. 140). Results showed faculty of color were subject to stereotyped-based expectations due to the students' lack of exposure to diversity (Reid, 2010). Reid (2010) concluded, ". . . racial minority faculty, particularly Black faculty, were evaluated more negatively than White faculty in terms of Overall Quality, Helpfulness, and Clarity, but were rated higher in Easiness" (p. 145). In another study, Perry et al. (2014) collected narratives from three black faculty members who explained their experiences with SET comments. Faculty recounted examples of negative responses, personal attacks, and challenges of authority by students, which they perceived as racially motivated (Perry et al., 2014). Perry et al. (2014) concluded, "It is likely that for many, the role that the instructor's race plays in student evaluations forces many African American faculty members to question and second guess themselves and their abilities" (p. 34).

In a similar study, Wallace et al. (2019) examined student comments submitted via an online survey by 46 members of the American Political Science Association. Respondents self-identified their racial backgrounds as "White (62.8%), Black (31.4%), Asian and Latino/a (2.8%), respectively" (p. 6). Wallace et al. (2019) found:

> The comments, assigned to women and faculty of color, tend to suggest overt and stereotypical connotations and tend be more derogatory and damaging which is consistent with the literature. Women and faculty of color were more likely to receive comments about their specific attributes such as likability or personality traits; appearance, mannerisms, languages style and demeanor. Additionally, the sample reveal some bias about perceived competence and professionalism (e.g. organizational skills, accessibility, support etc.) for women and faculty of color, as well as, elements of retaliation which can be seen in negative specific or complex comments. (p. 9)

Research regarding the impact of racial bias on the SET has been limited; however, colleges and universities are beginning to recognize the influence of implicit bias on SET results (Bavishi et al., 2010). Reinsch et al. (2020) stated, "Professors of color have published poignant accounts of harshly negative student evaluations. The few empirical studies examining instructor race and student ratings confirm that minority faculty receive significantly lower evaluations than their White colleagues" (p. 125). According to Reinsch et al. (2020), further research is needed to improve the retention of faculty of color and to reduce employment discrimination cases.

Gender

Gender has been identified in several studies as a factor in the results of the SET (Fan et al., 2019; MacNell, Driscoll, & Hunt, 2015; Reid, 2010). However, the degree of

influence gender has on SET results has been inconclusive due to the difficulty of separating other factors such as personality, experience, teaching style, and subject material (MacNell et al., 2015). Wallace et al. (2019) stated the perception that college teaching is a male profession may contribute to the low ratings received by female instructors.

Wallace et al. (2019) found students rate female instructors lower than male instructors on academic competency, professionalism, and organizational skills. Women of color experience a double-negative impact from SETs due to gender and race (Bavishi et al., 2010). Women of color are held to stereotypical status and have been evaluated by students as hostile, uncaring, and lacking interpersonal skills (Wallace et al., 2019). Female faculty of color are more likely to experience questioning of authority by students and colleagues (Sprague & Massoni, 2005; Wallace et al., 2019).

Sprague and Massoni (2005) suggested perceived gender roles impact the SET for both men and women; for example, men are expected to be entertaining and funny while women are expected to be caring and nurturing. Faculty who do not meet stereotypical gender role expectations receive harsh criticism on the SET (Sprague & Massoni, 2005). Sprague and Massoni (2005) stated, "The worst women teachers are sometimes explicitly indicted for being bad women through the use of words like bitch or witch" (p. 791). Reinsch et al. (2020) concluded, "The contradictory nature of the student comments on evaluations of minority faculty, the high levels of expressed hostility, and the occasional direct references to gender or race raise troubling questions about the role of bias in these assessments" (p. 126).

Research reveals a different criterion is used to evaluate female instructors in the areas of appearance, knowledge, personality, and overall competence (Mitchell & Martin,

2018). Reid (2010) noted, "For example, whereas a male faculty member can demonstrate competence and be unfriendly toward students and still be considered intellectually competent, a female faculty member must demonstrate competence and friendliness to be judged as intellectually competent" (p. 138). In an experiment conducted by MacNell et al. (2015), the gender of the instructors was falsified by posting a photograph and a gender-specific name to lead students to believe their instructors were male or female. Each instructor taught one course under his or her own identity and a second course under the false identity (MacNell et al., 2015). The results of the study revealed the ratings for the perceived male instructors were statistically significantly higher than those of the perceived female instructors (MacNell et al., 2015). Wallace et al. (2019) reported negative comments about white males were fewer in number and were focused on perceived course difficulties such as too much reading and too many assignments. Administrators viewed those comments positively because they indicated course rigor; however, the same comments about females were viewed by administrators as negative, indicating dissatisfied students (Wallace et al., 2019).

In a study conducted by Peterson, Biederman, Anderson, Ditonto, and Roe (2019), some students were given a standard SET survey while other students were given a SET with anti-bias language "intended to reduce gender bias" (p. 8). Peterson et al. (2019) concluded:

> The success of the anti-bias language, which make specific allusion to the unconscious and unintentional nature of biases, may be suggestive that the students' biases are implicit. It is also plausible that the intervention may have mitigated the use of more explicit gender bias. Regardless, the results do suggest that this intervention improved the SET scores for the female faculty. (p. 8)

Gender bias is a greater concern for women teaching in areas traditionally viewed as masculine fields of study (Mitchell & Martin, 2018). Reid (2010) discovered women teaching in disciplines such as physics receive lower SET ratings than in traditional female disciplines such as English. Owen (2019) added, "The bias in student evaluations is strongest in fields in which certain faculty members are underrepresented. For female faculty members, those are the male-dominated fields, such as many STEM fields and economics" (p. 3).

Rosen (2017) concluded women are simply at a disadvantage because their SET ratings are lower than men in every academic discipline. Sauermann, Mengel, and Zölitz (2019) stated, "Although the bias becomes somewhat smaller with higher course grades, students across the whole distribution make significantly worse evaluations when their instructors are female (18%-21% of a standard deviation)" (p. 31). McPherson et al. (2009) suggested departments that rank instructors according to SET scores should consider adjusting rankings to account for uncontrollable factors such as gender. According to Boring et al. (2016):

> We therefore conclude that SET primarily do not measure teaching effectiveness, that they are strongly and non-uniformly biased by factors including the genders of the instructor and student, that they disadvantage female instructors, and that it is impossible to adjust for these biases. (p. 2)

Research, according to Rosen (2017), has indicated acknowledgment of gender bias must be considered by administrators to ensure equity and fairness when interpreting SET results.

Age and Physical Attractiveness

Age and physical attractiveness are not related to teaching quality and should not impact the SET; however, Stalnaker and Remmers (1928) included personal appearance as an item on the Purdue Rating Sale to assess teaching effectiveness. McPherson and Jewell (2007) and Reints (2018) conducted studies and found age and physical appearance impact student responses on the SET. Prasad et al. (2017) stated, "Comments made as part of evaluations of female faculty are frequently unprofessional and/or of a personal nature, emphasizing such things as dress style and physical appearance" (p. 1368).

The online instructor review website RateMyProfessors.com has been criticized for rating professors based on physical attractiveness (Flaherty, 2017). Flaherty (2018) stated male and female instructors feel their work is not measured on the basis of teaching effectiveness due to the "hotness" rating question (p. 2). Flaherty (2018) added a tweet posted by one professor who expressed concern that male instructors were being sent a message that "female students wanted to sleep with them" (p. 2). Recent changes have occurred with the rating system of RateMyProfessors.com because of a social media campaign demanding the elimination of the chili pepper, which indicates the physical attractiveness of faculty (Reints, 2018). Reints (2018) added women have experienced disparaging remarks about their physical appearance, and in the age of the #Metoo movement, the elimination of the chili pepper was necessary.

Age is another instructor characteristic that has impacted the validity of the SET (Prasad et al., 2017). Arbuckle and Williams (2003) had students listen to and evaluate a recorded lecture of a gender- and age-neutral voice while viewing a gender- and age-neutral stick figure. The evaluation form included gender and age variations, so some

students were told the lecturer was young and male, while others were told the lecturer was some other variation (Arbuckle & Williams, 2003). All students viewed and listened to the same lecturer; however, students rated the young male lecturer higher than the older male, younger female, or older female for enthusiasm, voice tone, and interest in the subject (Arbuckle & Williams, 2003).

In a similar study conducted by Doubleday and Lee (2016), dental students watched a video narrated by one of four instructors, either a young man, young woman, older man, or older woman. Each instructor used an identical script, with the only difference being the individual presenting; ratings for the older female voice were significantly lower for most items on the questionnaire completed by the dental students (Doubleday & Lee, 2016). McPherson and Jewell (2007) found in their research that SET scores decrease with the age of the faculty member even though students indicate teaching experience is an important factor for effective teaching. According to McPherson and Jewell (2007), "One explanation for this finding may be a correlation between an instructor's age and his or her perceived "attractiveness," at least in terms of how students judge this quality" (p. 876).

English Proficiency

The impact of English proficiency on the SET is not a new concern for researchers. In an early review of SET studies, Kulik and Kulik (1974) stated, "What especially characterizes the highly rated teacher is verbal fluency and communication ability. The highly rated teacher seems cultured and sophisticated, expressive and enthusiastic" (p. 56). Reinsch et al. (2020) explained there are more opinion questions than behavioral questions on the SET, which can lead to bias in the responses. Specifically, Reinsch et al. (2020) added, "The accent issue is most problematic in 'the

instructor communicated course ideas in a clear and understandable manner.' Though most accents are perfectly understandable, they may trigger implicit bias" (p. 123). Schmidt (2017) explained colleges face a challenge determining if student concerns about a foreign accent are educational concerns or a sign of bias.

In a summary of key findings regarding SET bias, Huston (2006) noted non-native English speakers receive lower course evaluation ratings than native English speakers, and non-native English-speaking males are ranked lower than non-native English-speaking females. In a study conducted by Fan et al. (2019), SET data from an Australian university were examined to determine the impact of race, gender, and language background on surveys over a seven-year period. Results indicated a statistically significant bias against faculty with non-English language backgrounds, and Fan et al. (2019) recommended universities employ underrepresented groups to ensure students have more exposure to diversity, which will benefit society and perhaps reduce bias.

Summary

Reviewed in this chapter were the history of the SET, the theoretical framework of implicit bias, and the strengths and concerns of the SET. The impact of marginalization, race, ethnicity, gender, age, physical attractiveness, and English proficiency were also examined in this chapter. While the SET is an economical and common method to assess teaching effectiveness, instructor characteristics such as race, ethnicity, English proficiency, gender, age, and physical attractiveness may impact the results (Boring et al., 2016; Wallace et al., 2019).

In the next chapter, the methodology of this study is described. An overview of the problem, purpose, research design, and research questions is presented. The

population and sample, instrumentation, and data collection processes are explained. The final section contains information about the analysis of data.

Chapter Three: Methodology

In this chapter, the research methodology is discussed. Quantitative data from student and faculty surveys were collected to determine the influence of implicit bias on the outcomes of student evaluations of teaching (SET) at one Missouri community college. An explanation of the problem and purpose is provided, and the research questions are restated. This chapter also includes a description of the research design, the population and sample, the instrumentation, and the processes for data collection and analysis.

Problem and Purpose Overview

As stated in Chapter One, the SET is often used by administrators to determine the teaching effectiveness of faculty, which impacts hiring, promotion, tenure, and other employment decisions (Calkins & Micari, 2010; Feistauer & Richter, 2018; McPherson & Jewell, 2007; Rosen, 2017). The outcomes of the SET can change the course of a faculty member's career and life, so determining the validity of the SET is vital to faculty and administrators (Wallace et al., 2019). There were two primary goals for this study. The first was to determine to what extent students evaluate faculty teaching effectiveness based upon instructor characteristics. The second was to determine to what degree faculty perceived they were evaluated by a criterion based on instructor characteristics, not on their actual teaching. Data from a student and faculty survey were collected during the fall 2019 semester from students enrolled in courses within the Communication and World Languages department at the community college.

Research questions. The following questions guided the research in this study:

1. What percentage of students utilize a different criterion to evaluate faculty from marginalized groups than faculty from non-marginalized groups?

2. Based on the opinions of students and faculty, to what extent does implicit bias influence the outcome of student evaluations of teaching (SET)?

3. Based on the opinions of students and faculty, to what extent do instructor characteristics predict the outcome of student evaluations of teaching (SET)?

Research Design

A quantitative approach was the most appropriate method for this study since this method allowed for the examination of numerical data (Creswell & Creswell, 2017). The data were obtained from two Likert-type surveys, which were adapted from a study conducted by Dev and Qayyum (2017). In that study, student perceptions toward SET evaluations were examined to determine the validity and reliability of responses along with the influence of various factors such as instructor nationality, age, and gender (Dev & Qayyum, 2017). The adaptation for this study included a faculty survey, which mirrored the content of the student survey from Dev and Qayyum's (2017) original survey. Additional statements were added to address other variables in this study that were not relevant to Dev and Qayyum's (2017) study.

Population and Sample

The Missouri community college examined in this study noted a population of approximately 12,000 students, as reported through the Integrated Postsecondary Education Data System (NCES, 2019). The student population for this study included a maximum of 2,000 students enrolled in Spanish, French, American Sign Language, and Communication courses from all sites of this college (M. Tollett, personal communication, September 14, 2018). The faculty population of the college system was comprised of approximately 1,100 instructors, of which 60 faculty members were within the Communication and World Languages department from all sites of this college (M.

Tollett, personal communication, September 14, 2018). There were no sampling methods used for this study. The number of students and faculty who chose to respond to the survey determined the sample size. Cook, Heath, and Thompson (2000) indicated an average response rate of 25%-35% might be expected, which would yield a sample of approximately 20 faculty members and 400 students for this study. The participation rate for faculty was above average with 34 faculty participants; however, student participation was below average with only 245 students.

Instrumentation

According to Fraenkel et al. (2019), "A survey is a tool for collecting information from a sample population by asking questions to obtain opinions about an issue or topic" (p. 393). The instruments for this study included two survey adaptations of Dev and Qayyum's (2017) survey (see Appendices A and B). Permission to modify Dev and Qayyum's survey instrument was granted by Dr. Smitha Dev (see Appendix C). The surveys were administered through an online system hosted by Qualtrics.

The first section of each survey was composed of demographic questions requesting the age, gender, and race of the respondents. The second section contained 16 statements requiring responses of *strongly agree, agree, neutral, disagree,* or *strongly disagree* to gather data about the respondents' perceptions of the SET.

Reliability. The student perception tool administered to collect data for this research was reliable. Dev and Qayyum (2017) stated, "The test-retest reliability was established by re-administrating the test and computing reliability coefficient for total test. The reliability coefficient is reported to be +0.92 on a sample of 50 students with the time interval of 2 weeks" (p. 157). Since the original student perception tool was amended for this study, a pilot version of the survey was administered to faculty

volunteers recruited by the researcher. The faculty survey, created based on the student perception tool, was administered to the same volunteers for their comments.

Validity. The goal of the validity of a quantitative study is to minimize errors and biases, and this can be alleviated by choosing the survey method that is most precise and accurate for the specific purposes (Fink, 2016). Dev and Qayyum (2017) stated, "Validity of the scale was established by correlating the scores of the scale with the scale of Student Perception Scale used in AUS, and validity coefficients were found to be +0.84 and +0.79, respectively" (p. 157).

Data Collection

To collect the necessary data information, approval from the Institutional Review Board (IRB) was obtained (see Appendix D), and permission was obtained from the academic dean of the participating college. Permission was also obtained from the academic dean of the participating college. Data from students and faculty were gathered from two similar surveys. Students currently enrolled at the time of this study within the Department of Communication and World Languages were asked to complete an electronic survey distributed through Qualtrics. Each student received a letter of participation (see Appendix E) and an informed consent form (see Appendix G) to read and determine if he or she chose to complete the survey. Each faculty member also received a letter of participation (see Appendix F) and informed consent form (see Appendix H). Faculty members were asked to complete a similar electronic survey distributed through Qualtrics following the same process for collection. The identities of the participants were protected because identifying information was not collected in the survey process. To increase responses, periodic reminders were sent to students and faculty requesting participation.

Data Analysis

The raw data were downloaded into a file, and the contents were reported by Qualtrics. The demographic information was presented using percentages. Data collected from both surveys were analyzed by applying measures of central tendency, specifically "a simple percentage analysis" (Dev & Qayyum, 2017, p. 160). Extracted data were disaggregated according to the variables of student and instructor gender, race, and age. Survey statements were grouped into three categories: statements that identified different criteria used to evaluate faculty from marginalized groups, statements that identified the extent of implicit bias on the outcome of the SET, and statements that identified the impact of instructor characteristics.

The first three items of the survey addressed the demographics of the participants with questions about age, race, and gender. Survey statements 1-4 addressed the perceptions of faculty and students regarding the importance of the SET and the SET's ability to determine teaching effectiveness. Student survey statements six, eight, nine, and 15 addressed research question one. Student and faculty survey statements five, seven, 10, and 16 addressed research question two. Finally, student and faculty survey statements 11, 12, and 13 addressed research question three.

Ethical Considerations

Fraenkel et al. (2019) stated bias "occurs when the design of a study systematically favors certain outcomes" (p. G-1). Research bias was possible in this study for the following reasons. The researcher acknowledged an association with participants as the Department Chair of Communication and World Languages at the participating college. There was a possibility student participants were enrolled in classes taught by the researcher. The researcher was the supervisor of the faculty

participants. The identities of student and faculty participants were protected through the survey process because no identifying information was collected.

Summary

In this chapter, the problem and purpose of the study, including three research questions that guided the study, were presented. In addition, the population, sample, and the process of data collection and analysis were described. Ethical considerations were stated.

In Chapter Four, data collected from the student and faculty surveys are presented and analyzed. The demographic data of student and faculty respondents are described. Finally, the results of this study are presented.

Chapter Four: Analysis of Data

Student evaluations of teaching (SET) are used by post-secondary institutions to quantify instructor performance (Uttl et al., 2017). The SET was created in the 1920s by Remmers from Purdue University (Stalnaker & Remmers, 1928). Since its inception, the SET has been troublesome to some faculty and administrators due to its potential susceptibility to bias and manipulation (Bonitz, 2011; Ray, 2018).

As reviewed in Chapter Two, numerous researchers have demonstrated the impact implicit bias can have on the outcomes of the SET (Dev & Qayyum, 2017). The potential of lower ratings for instructors from marginalized groups can result in negative outcomes for those instructors regarding personnel decisions such as hiring, promotion, tenure, and termination (Basow et al., 2013). In this chapter, quantitative results from the study are discussed.

The primary purpose of this study was to explore how instructor characteristics such as ethnicity, race, age, gender, accent, and personality influence student responses on the SET at a Missouri community college. Faculty in the Communication and World Languages department were asked to complete a survey comprised of 16 Likert-type statements to provide their perceptions of the SET. Students enrolled in these courses at the time of the study were asked to provide responses to a similar survey comprised of 16 Likert-type statements. Each of the possible responses were assigned a value with *Strongly Agree* receiving a score of five and *Strongly Disagree* receiving a score of one. A mean score and standard deviation were calculated for the responses.

Demographic Data

On the student survey, three questions regarding demographics were asked to determine gender, student race, and student age. On the faculty survey, the same demographic information was requested.

Data Analysis

Descriptive analysis was used to analyze data obtained from the surveys administered to student and faculty respondents.

Student demographic statements. Respondents provided their ages by selecting one of four options: 18-25, 26-40, 41-55, or 56 and older. The majority (80%) of respondents identified as traditional college age of 18-25, while 12.24% were in the 26-40 range, 6.12% were in the 41-55 age range, and 1.63% were 56 or older. Respondents were also asked to provide their gender. Most respondents were female (67.62%), 32.28% were male, and 0.1% did not identify their gender. The final demographic prompt was about race. Most respondents were White (86.12%), followed by Hispanic/Latino (6.12%), Asian/Pacific Islander (2.6%), Native American/American Indian (2.6%), Black/African American (2.04%), and Other (2.86%). Student demographics are displayed in Table 1.

Table 1

Student Respondent Demographics

Demographic Category	Option	*f*	%
Age	18-25	196	80
	26-40	30	12.24
	41-55	15	6.12
	56+	4	1.63
Gender	Female	165	67.62
	Male	79	32.28
Race	White	211	86.12
	Black/African American	5	2.04
	Hispanic/Latino	15	6.12
	Asian/Pacific Islander	5	2.04
	Native American/American Indian	2	.82
	Other	7	2.86

Note. $N = 245$, *f* = frequency, % = percentage.

Student survey statement one. *By evaluating my professors, I am actually helping them improve their teaching.* For this survey statement, 38.37% of respondents strongly agreed, 40.41% agreed, 14.29% selected neutral, 4.49% disagreed, and 2.45% strongly disagreed. Male participants and female participants responded similarly with 73.42% of males and 81.21% of females indicating strongly agree or agree. Another 12.73% of females and 17.72% of males selected neutral, while 6.06% of females selected disagree or strongly disagree compared to 8.86% of males.

When disaggregating the data by age range, most respondents selected strongly agree or agree to this statement (18-25 = 79.09%, 26-40 = 86.67%, 41-55 = 80%); however, none of the respondents over the age of 55 indicated strongly agree or agree as a response. Most respondents over the age of 55 (75%) indicated neutral compared to only 11.26% of respondents age 18-55. Disagree or strongly disagree was selected at the

following rates: 18-25 (7.14%), 41-55 (13.34%), and 56 and older (25%). No respondents age 26-40 indicated disagree or strongly disagree.

When disaggregating the data by race, most respondents selected strongly agree or agree with the following frequency: White (78.20%), Black/African American (100%), Hispanic/Latino (80%), Asian/Pacific Islander (80%), Native American/American Indian (100%), and Other (71.43%). Only 8.05% of those identifying as White selected disagree or strongly disagree. Based on the results of all respondents, a mean score with regard to the SET improving teaching was calculated at 1.92 with a standard deviation of 0.96.

Student survey statement two. *The course evaluation form is adequate to evaluate my professors.* In response to survey statement two, 28.16% of respondents strongly agreed, 48.98% agreed, 15.51% selected neutral, 6.94% disagreed, and 0.41% strongly disagreed. Female respondents (84.85%) indicated strongly agree or agree at a greater rate than male respondents (60.76%). More male respondents selected disagree or strongly disagree (13.93%) when compared to females (4.24%).

When disaggregating the data by age range, most respondents selected strongly agree or agree as their response to this statement (18-25 = 79.59%, 26-40 = 70%, 41-55 = 73.33%); however, only 25% of respondents 56 or older selected strongly agree or agree. Respondents selected neutral at the following frequencies: 18-25 (15.82%), 26-40 (10%), 41-55 (13.33%), and 56 and older (50%). Those selecting disagree were disaggregated as follows: 18-25 (4.59%), 26-40 (20%), 41-55 (6.67%), and 56 and older (25%). Only respondents age 41-55 selected strongly disagree (6.67%).

When disaggregating the data by race, most respondents indicated strongly agree or agree with the following frequency: White (75.83%), Hispanic/Latino (93.33%), Asian/Pacific Islander (80%), Black/African American (100%), Native American/

American Indian (100%), and Other (57.14%). Neutral was selected at the following rates: White (16.11%), Hispanic/Latino (6.67%), Asian/Pacific Islander (20%), and Other (28.57%). Only 8.05% of those identifying as White and 14.29% of those identifying as Other indicated disagree or strongly disagree. Based on the results of all respondents, a mean score with regard to the SET's ability to evaluate professor performance was calculated at 2.02 with a standard deviation of 0.087.

Student survey statement three. *Students should take student evaluations of teaching seriously.* For this survey item, 60% of respondents strongly agreed, 34.29% agreed, and 5.71% selected neutral. No respondents indicated they disagreed. Nearly all female respondents (98.18%) selected strongly agree or agree, compared to male respondents (86.07%). Neutral was selected by female respondents (1.82%) less frequently than by males (13.92%).

When disaggregating by age range, most respondents selected strongly agree or disagree with the following frequencies: 18-25 (93.87%), 26-40 (96.67%), 41-55 (93.34%), and 56 and older (100%). When disaggregating the data by race, more than 90% of those identifying as White, Black/African American, Native American/American Indian, and Other selected strongly agree or agree. Only respondents identifying as Hispanic/Latino and as Asian/Pacific Islander selected neutral (20%). Based on the results of all respondents, a mean score with regard to student perception of the seriousness of the SET was calculated at 1.46 with a standard deviation of 0.60.

Student survey statement four. *I read and understand each statement before I rate it.* Again, over 90% of respondents agreed the SET should be taken seriously by reading and understanding each statement. For this survey item, 74.69% strongly agreed, 24.08% agreed, 0.82% selected neutral, and 0.41% disagreed. Male respondents and

female respondents had similar responses, with 97.47% of male respondents and 99.39% of female respondents indicating strongly agree or agree.

When disaggregating the data by age, most respondents selected strongly agree or agree with the following frequencies: 18-25 (98.98%), 26-40 (100%), 41-55 (93.33%), and 56 and over (100%). Neutral was only selected by respondents age 18-25 (1.02%), and disagree was only selected by respondents age 41-55 (6.67%). When disaggregating the data by race, most respondents indicated strongly agree or agree as follows: White (99.06%), Hispanic/Latino (100%), Asian/Pacific Islander (100%), Black/African American (100%), Native American/American Indian (100%), and Other (100%). Neutral was only selected by white respondents (0.94%), and disagree was only selected by respondents identifying as Asian/Pacific Islander (20%). Based on the results of all respondents, a mean score with regard to the importance of reading and understanding SET statements was calculated at 1.27 with a standard deviation of 0.49.

Student survey statement five. *I prefer taking courses from male instructors.* For survey statement five, most student respondents (76.64%) selected neutral, indicating no opinion regarding the gender of instructors. However, 10.63% of respondents selected strongly agree or agree, indicating some preference for male instructors. Another 12.71% of respondents selected disagree or strongly disagree, indicating no preference for male instructors.

When disaggregating the data by gender, most respondents selected neutral including 76.22% of females and 77.22% of males. Respondents selected strongly agree or agree as a response with the following frequency: female (11.59%) and male (8.86%). Another 13.90% of male respondents and 12.22% of female respondents selected disagree or strongly disagree.

When disaggregating the data by age, most respondents selected neutral at the following frequencies: 18-25 (76.64%), 26-40 (76.67%), 41-55 (86.67%), and 56 and older (50%). Strongly agree and agree were indicated as follows: 18-25 (10.66%), 26-40 (16.67%), and 56 and older (25%). No respondents age 41-55 selected strongly agree or agree. Respondents selected disagree or strongly disagree at the following rates: 18-25 (12.71%), 26-40 (6.66%), 41-55 (13.34%), and 56 and older (25%).

When disaggregating the data by race, most respondents selected neutral with the following frequencies: White (75.24%), Black/ African American (80%), Hispanic/Latino (86.67%), Asian/Pacific Islander (100%), Native American/American Indian (100%) and Other (71.43%). Strongly agree and agree were indicated by 10.95% of White respondents, 30% of Black/African American respondents, and 13.33% of Hispanic/Latino respondents. No respondents of other races selected strongly agree or agree. Disagree or strongly disagree was indicated by 13.81% of White participants and 28.58% of those who identified as Other. No respondents of other races selected disagree or strongly disagree. Based on the results of all respondents, a mean score with regard to student preference for male instructors was calculated at 3.03 with a standard deviation of 0.61. The results for student statement five responses by gender are displayed in Table 2.

Table 2

Summary of Student Statement Five Responses by Gender

Gender	Student Response	*f*	%
Male	Strongly Agree or Agree	7	8.86
	Neutral	61	77.22
	Disagree or Strongly Disagree	11	13.90
Female	Strongly Agree or Agree	19	11.59
	Neutral	125	76.22
	Disagree or Strongly Disagree	20	12.20

Note. $N = 243$, *f* = frequency, % = percentage.

Student survey statement six. *I prefer taking courses from female instructors.* Similar to survey statement five, most student respondents (78.78%) selected neutral, indicating no opinion regarding the gender of instructors. However, 14.29% of respondents selected strongly agree or agree, indicating some preference for female instructors. Only 6.94% of respondents strongly disagreed or disagreed with the statement.

When disaggregating the data by gender, 78.48% of males and 78.79% of females selected neutral. Strongly agree and agree were selected by 17.73% of males and 12.72% of females, while disagree or strongly disagree were selected by 8.48% of females and 3.80% of males. When disaggregating the data by age, most respondents selected neutral with the following frequencies: 18-25 (80.61%), 26-40 (70%), 41-55 (80%), and 56 and older (50%). Strongly agree and agree were indicated with the following percentages: 18-25 (14.29%), 26-40 (20%), and 41-55 (6.67%). No respondents 56 or older selected strongly agree or agree. Strongly disagree and disagree were indicated as follows: 18-25 (5.1%), 26-40 (10%), 41-55 (13.34%), and 56 and older (50%).

When disaggregating the data by race, most respondents selected neutral as follows: White (78.67%), Black/African American (60%), Hispanic/Latino (80%), Asian/Pacific Islander (100%), Native American/American Indian (100%), and Other (77.78%). Strongly agree and agree were indicated as the responses by 14.22% of White participants, 40% of Black/African American participants, and 20% of Hispanic/Latino participants. No respondents of other races selected strongly agree or agree. Disagree or strongly disagree were indicated with the following percentages: White (7.11%) and Other (22.22%). No respondents of other races selected disagree or strongly disagree. Based on the results of all respondents, a mean score with regard to student preference for female instructors was calculated at 2.94 with a standard deviation of 0.61. The results for student statement six responses by gender are displayed in Table 3.

Table 3

Summary of Student Statement Six Responses by Gender

Gender	Student Response	*f*	%
Male	Strongly Agree or Agree	14	17.73
	Neutral	62	78.48
	Disagree or Strongly Disagree	3	3.80
Female	Strongly Agree or Agree	21	12.72
	Neutral	130	78.79
	Disagree or Strongly Disagree	14	3.03

Note. $N = 245$, f = frequency, % = percentage.

Student survey statement seven. *I prefer taking courses from young and enthusiastic instructors.* For this survey item, 6.97% of respondents strongly agreed, 27.46% agreed, 59.43% selected neutral, 5.33% disagreed, and only 0.82% strongly disagreed. Male respondents indicated strongly agree or agree 40.5% of the time compared to female respondents with 31.71%. Of the female respondents, 62.2%

selected neutral compared to 53% of male respondents. Disagree or strongly disagree was selected as a response by 6.1% of females and 6.33% of males.

When disaggregating the data by age range, most respondents selected neutral with the following frequencies: 18-25 (55.38%), 26-40 (70%), 41-55 (86.67%), and 56 and older (75%). Strongly agree and agree were indicated by participants as follows: 18-25 (39.49%), 26-40 (20%), and 41-55 (6.67%). No respondents in the age group 56 or older selected strongly agree or agree. Strongly disagree and disagree were indicated by the following percentages: 18-25 (5.13%), 26-40 (10%), 41-55 (6.67%), and 56 and older (25%).

When disaggregating the data by race, strongly agree and agree were indicated as follows: White (35.72%), Black/African American (80%), Hispanic/ Latino (20%), Asian/Pacific Islander (20%), and Native American/American Indian (50%). No respondents in the Other category selected strongly agree or agree. The respondents who selected neutral included the following: White (59.05%), Black/African American (20%), Hispanic/Latino (73.33%), Asian/Pacific Islander (80%), Native American/American Indian (50%), and Other (57.14%). Disagree or strongly disagree was indicated as follows: White (5.24%), Hispanic/ Latino (6.67%), and Other (42.86%). No respondents of Other races selected disagree or strongly disagree. Based on the results of all respondents, a mean score with regard to student preference for young and enthusiastic instructors was calculated at 2.66 with a standard deviation of 0.72.

Student survey statement eight. *I prefer taking courses from older and experienced instructors.* Similar to survey statement seven, most students selected neutral (59.18%). The remaining students responded to this statement as follows: 7.76% strongly agreed, 28.16% agreed, and only 4.9% selected disagree or strongly disagree.

When disaggregating the data by gender, strongly agree or agree were selected in response to this statement by 44.30% of males and 32.12% of females. Neutral was selected as follows: female (61.82%) and male (53.16%). Disagree or strongly disagree was the response of 6.06% of females and 2.53% of males.

When disaggregating the data by age range, respondents selected strongly agree or agree to this statement with the following frequencies: 18-25 (36.73%), 26-40 (43.34%), and 41-55 (20%). No respondents 56 and older selected strongly agree or agree. Neutral was indicated as the response with the following percentages: 18-25 (57.65%), 26-40 (53.33%), 41-55 (80%), and 56 and older (100%). Strongly disagree and agree were indicated as follows: 18-25 (5.61%) and 26-40 (3.33%). No respondents age 41 and older indicated disagree or strongly disagree.

When disaggregating the data by race, most respondents selected neutral as follows: White (60.66%), Black/African American (60%), Hispanic/Latino (53.33%), Asian/Pacific Islander (60%), and Other (42.86%). No Native American/American Indian respondents selected neutral as a response to this statement. Strongly agree and agree were indicated with the following frequencies: White (35.07%), Black/African American (40%), Hispanic/ Latino (33.34%), Asian/Pacific Islander (40%), Native American/American Indian (100%), and Other (42.86%). Disagree or strongly disagree was indicated as by 4.26% of White participants, 13.33% of Hispanic/Latino participants, and 14.29% of those who identified as Other. Based on the results of all respondents, a mean score with regard to student preference for older and experienced instructors was calculated at 2.62 with a standard deviation of 0.71.

Student survey statement nine. *I am comfortable taking courses from instructors who speak English as their second language.* For survey statement nine, half of the

students (50.2%) felt comfortable taking courses from instructors who speak English as their second language. Respondents selected strongly agree (19.18%), agree (31.02%), neutral (26.94%), disagree (19.59%), and strongly disagree (3.27%). When disaggregating the data by gender, strongly agree or agree was selected by 40.51% of males and 54.55% of females. Neutral was selected as a response as follows: female (24.24%) and male (32.91%). Disagree or strongly disagree was selected as a response by 21.21% of female participants and 26.58% of male participants.

When disaggregating the data by age range, respondents selected strongly agree or agree as follows: 18-25 (51.02%), 26-40 (46.66%), 41-55 (40%), and 56 and older (75%). Neutral was indicated as the response with the following frequencies: 18-25 (24.49%), 26-40 (33.33%), 41-55 (46.67%), and 56 and older (25%). Strongly disagree and disagree were indicated by 24.49% of those age 18-25, 20% of those age 26-40, and 13.33% of those age 41-55. No respondents age 56 and older indicated disagree or strongly disagree.

When disaggregating the data by race, strongly agree and agree were indicated as follows: White (47.39%), Black/African American (60%), Hispanic/Latino (73.33%), Asian/Pacific Islander (60%), and Native American/American Indian (100%). Respondents selected neutral with the following frequencies: White (26.54%), Black/African American (40%), Hispanic/Latino (26.67%), Asian/Pacific Islander (20%), and Other (42.86%). No Native American/American Indian respondents selected neutral. Disagree or strongly disagree was only indicated as the response to this statement by respondents who identified as White (26.06%). Based on the results of all respondents, a mean score with regard to student comfort with instructors who speak English as their

second language was calculated at 2.57 with a standard deviation of 1.10. The results for student statement nine responses by race are displayed in Table 4.

Table 4

Summary of Student Statement Nine Responses by Race

Student Race	Student Response	*f*	%
White	Strongly Agree or Agree	100	47.39
	Neutral	56	26.54
	Disagree or Strongly Disagree	55	26.06
Non-White	Strongly Agree or Agree	23	67.65
	Neutral	10	29.41
	Disagree or Strongly Disagree	1	2.94

Note. $N = 245$, *f* = frequency, % = percentage. Non-White = Hispanic/Latino, Asian/Pacific Islander, Native American/American Indian, Black/African American, and/or Other.

Student survey statement 10. *I am more comfortable taking courses from instructors who speak English as their first language.* For this survey question, 27.76% strongly agreed, 31.43% agreed, 35.92% selected neutral, 35.92% disagreed, and 1.22% strongly disagreed. Strongly agree or agree was indicated as the response to this statement by 64.56% of males and 56.97% of females. Respondents indicating neutral included 35.44% of males and 35.76% of females. No male respondents selected disagree or strongly disagree, while 8.86% of female respondents did so.

When disaggregating the data by age, respondents selected strongly agree or agree with the following percentages: 18-25 (62.24%), 26-40 (43.34%), 41-55 (53.33%), and 56 and over (50%). Neutral was indicated as follows: 18-25 (35.71%), 26-40 (36.67%), 41-55 (40%), and 56 and older (25%). Disagree or strongly disagree was indicated with the following frequencies: 18-25 (2.04%), 26-40 (20%), 41-55 (6.67%), and 56 and over (25%).

When disaggregating the data by race, respondents selected strongly agree or agree as follows: White (61.13%), Black/African American (80%), Hispanic/Latino (33.34%), Asian/Pacific Islander (60%), Native American/American Indian (50%), and Other (44.44%). Neutral was selected with the following percentages: White (35.07%), Black/African American (20%), Hispanic/Latino (60%), Asian/Pacific Islander (20%), Native American/American Indian (50%), and Other (33.33%). Disagree or strongly disagree was indicated as follows: White (3.79%), Hispanic/ Latino (6.67%), Asian/ Pacific Islander (20%), and Other (22.22%). Based on the results of all respondents, a mean score with regard to student preference for instructors who speak English as their first language was calculated at 2.19 with a standard deviation of 0.93.

Student survey statement 11. *When evaluating professors, I pay more attention to their personality (i.e., friendliness, leniency, looks, or dress).* For survey statement 11, respondents selected strongly agree (12.24%), agree (38.37%), neutral (25.31%), disagree (20%), and strongly disagree (4.08%). Male and female respondents had similar responses, with 51.90% of males indicating strongly agree or agree compared to 49.70% of the female respondents. Neutral was selected as the response to this statement by 25.32% of females and 25.45% of males. Disagree or strongly disagree was selected as follows: females (22.78%) and males (24.85%).

When disaggregating the data by age, respondents selected strongly agree and agree with the following frequencies: 18-25 (54.60%), 26-40 (40%), 41-55 (60%), and 56 and over (25%). Neutral was indicated as follows: 18-25 (24.49%), 26-40 (26.67%), 41-55 (33.33%), and 56 and older (25%). Disagree or strongly disagree was selected by respondents with the following percentages: 18-25 (20.92%), 26-40 (33.33%), 41-44 (40%), and 56 and older (50%).

When disaggregating the data by race, the following respondents selected strongly agree or agree: White (47.87%), Black/African American (100%), Hispanic/Latino (46.66%), Asian/Pacific Islander (100%), Native American/American Indian (100%), and Other (66.66%). Neutral was indicated as the response to this statement as follows: White (26.07%), Hispanic/Latino (33.33%), and Other (22.22%). No additional respondents selected neutral. Disagree or strongly disagree was indicated as the response by 26.07% of White participants, 53.33% of Hispanic/Latino participants, and 33.3% of those who identified as Other (33.33%). No additional respondents selected disagree or strongly disagree to this statement. Based on the results of all respondents, a mean score regarding the student's focus on an instructor's personality (i.e., friendliness, leniency, looks, or dress) was calculated at 2.65 with a standard deviation of 1.06. The results for student statement 11 responses are displayed in Table 5.

Table 5

Student Survey Statement 11 Responses

Student Demographics Category	Student Response	*f*	%
All Students	Strongly Agree or Agree	124	50.61
	Neutral	62	25.31
	Disagree or Strongly Disagree	59	24.08

Note. $N = 245$, f = frequency, % = percentage.

Student survey statement 12. *I prefer instructors who are assertive and provide clear expectations for the course.* Over 80% of students indicated a preference for assertive instructors who provide clear expectations. For this survey item, 38.52% strongly agreed, 47.13% agreed, and 11.89% selected neutral. Only 2.45% of respondents indicated they disagreed. Male respondents and female respondents had similar responses with 85.97% of males indicating strongly agree or agree and 84.81% of

females indicating strongly agree or agree. Another 11.59% of females and 12.66% of males selected neutral, and 2.44% of females selected disagree compared to 2.53% of males.

When disaggregating the data by age range, most respondents selected strongly agree or agree as follows: 18-25 (85.13%), 26-40 (90%), 41-55 (80%), and 56 and over (100%). Neutral was selected by 11.79% of those age 18-25, 10% of those age 26-40, and 40% of those age 41-55. No respondents age 56 or older indicated neutral as a response to this statement. Only respondents who identified as White (3.06%) selected disagree as a response to this statement.

When disaggregating the data by race, respondents selected strongly agree or agree with the following percentages: White (86.66%), Black/African American (100%), Hispanic/Latino (80%), Asian/Pacific Islander (80%), Native American/American Indian (100%), and Other (71.43%). Only White (2.38%) and Hispanic/Latino respondents (6.67%) selected disagree as a response to this statement. Based on the results of all respondents, a mean score with regard to student preference for assertive instructors was calculated at 1.78 with a standard deviation of 0.75.

Student survey statement 13. *I prefer instructors who are caring and concerned about me outside of the classroom.* Similar to survey statement 12, over 80% of student respondents indicated a preference for caring and concerned instructors. For this survey statement, 34.29% selected strongly agree, 48.57% selected agree, 14.69% selected neutral, and only 2.4% selected disagree. When disaggregating the data by gender, strongly agree or agree was selected as a response to this statement by 81.02% of males and 83.64% of females. Neutral was selected as follows: females (13.33%) and males (17.72%). Disagree was selected as a response by 3.03% of females and 1.27% of males.

When disaggregating the data by age range, most respondents selected strongly agree or agree with the following frequencies: 18-25 (84.18%), 26-40 (80%), 41-55 (73.33%), and 56 and over (75%). Neutral was indicated as the response as follows: 18-25 (14.80%), 26-40 (20%), and 41-55 (6.67%). No respondents age 56 and older indicated neutral as a response to this statement. Disagree was indicated as the response by 1.02% of those age 18-25, 20% of those age 41-55, and 25% of those age 56 and older. No respondents age 41-55 indicated disagree.

When disaggregating the data by race, most respondents selected strongly agree or agree. Specifically, those who agreed or strongly agreed were White (83.41%), Black/African American (100%), Hispanic/Latino (60%), Asian/Pacific Islander (100%), Native American/American Indian (100%), and Other (85.71%). Neutral was only selected as the response to this statement by White (15.17%) and Hispanic/Latino (26.67%) respondents. Disagree was indicated as the response to this statement by 1.42% of White participants, 13.33% of Hispanic/Latino participants, and 14.29% of the participants who identified as Other. Based on the results of all respondents, a mean score with regard to student preference for caring instructors was calculated at 1.85 with a standard deviation of 0.75.

Student survey statement 14. *I have taken a class from an instructor of a different race or ethnicity*. For survey statement 14, 56.79% of student respondents indicated they had taken a class from an instructor of a different race or ethnicity, and 43.21% of student respondents indicated they had not taken a class from an instructor of a different race or ethnicity. When disaggregating the data by gender, male respondents (51.28%) and female respondents (48.72%) indicated they had taken a class from an instructor of a different race or ethnicity. When disaggregating the data by age range, the

following respondents selected agree to this statement: 18-25 (52.82%), 26-40 (80%), 41-55 (64%), and 56 and over (50%).

When disaggregating the data by race, respondents selected agree with the following percentages: White (54.55%), Black/African American (40%), Hispanic/Latino (73.33%), Asian/Pacific Islander (100%), Native American/American Indian (50%), and Other (71.43%). Based on the results of all respondents, a mean score with regard to student exposure to an instructor from a different race or ethnicity was calculated at 1.43 with a standard deviation of 0.50%. The results for student statement 14 responses are displayed in Table 6.

Table 6

Student Survey Statement 14 Responses

Student Demographics Category	Student Response	*f*	%
All Students	Agree	138	56.79
	Disagree	105	43.21

Note. $N = 243$, f = frequency, % = percentage.

Student survey statement 15. *I am excited to take a class from an instructor from a different race or ethnicity than my own.* For this survey statement, 21.63% strongly agreed, 30.20% agreed, 45.31% selected neutral, 2.45% disagreed, and only 0.41% strongly disagreed. When disaggregating the data by gender, strongly agree or agree was selected as a response by 56.97% of females and 40.50% of males. Neutral was selected as a response by 40% of females and 56.96% of males. Disagree or strongly disagree was selected as follows: females (3.03%) and males (2.54%).

When disaggregating the data by age range, respondents selected strongly agree or agree with the following frequencies: 18-25 (52.04%), 26-40 (56.67%), 41-55 (40%), and 56 and older (50%). Neutral was indicated as follows: 18-25 (45.41%), 26-40 (40%),

41-55 (53.33%), and 56 and older (50%). Strongly disagree and disagree were indicated by 2.55% of those age 18-25, 3.33% of those age 26-40, and 6.67% of those age 41-55. No respondents age 56 and older indicated disagree or strongly disagree as a response to this statement.

When disaggregating the data by race, respondents selected strongly agree or agree as follows: White (48.34%), Black/African American (80%), Hispanic/Latino (73.33%), Asian/Pacific Islander (80%), Native American/American Indian (50%), and Other (66.66%). Neutral was indicated with the following percentages: White (48.34%), Hispanic/Latino (20%), Asian/Pacific Islander (20%), Native American/American Indian (50%), and Other (33.33%). Disagree or strongly disagree was indicated as the response to this statement by those identifying as White (3.31%). Based on the results of all respondents, a mean score with regard to student interest in taking a course from an instructor from a different race or ethnicity was calculated at 2.30 with a standard deviation of 0.85.

Student survey statement 16. *I am more comfortable with instructors who appear to be from my own racial or ethnic background.* For this survey statement, 7.76% of respondents strongly agreed, 11.43% agreed, 51.43% selected neutral, and 29.39% disagreed. When disaggregating the data by gender, strongly agree or agree was selected as a response by 17.72% of male and 20% of females. Neutral was selected as follows: female (47.88%) and male (58.23%). Disagree or strongly disagree was selected as a response by 32.12% of female participants and 24.05% of male participants.

When disaggregating the data by age range, respondents selected strongly agree or agree as follows: 18-25 (23.47%) and 41-55 (6.67%). No respondents age 26-40 or 56 and older selected strongly agree or agree as a response to this statement. Neutral was

indicated with the following percentages: 18-25 (52.04%), 26-40 (56.67%), 41-55 (33.33%), and 56 and older (50%). Strongly disagree and disagree were indicated as follows: 18-25 (24.49%), 26-40 (43.33%), 41-55 (60%), and 56 and older (50%).

When disaggregating the data by race, respondents selected strongly agree or agree with the following frequencies: White (18.01%), Black/African American (80%), and Hispanic/Latino (33.33%). No other races selected strongly agree or disagree. Neutral was selected as follows: White (53.55%), Black/African American (20%), Hispanic/Latino (40%), Asian/Pacific Islander (100%), and Native American/American Indian (50%). No respondents identifying as Other selected neutral as a response to this statement.

Disagree was indicated by the following: White (28.44%), Hispanic/Latino (26.67%), Native American/American Indian (50%), and Other (100%). No respondents identifying as Black/African American or Asian/Pacific Islander selected disagree as a response to this statement. Based on the results of all respondents, a mean score with regard to student preference for instructors who appear to be from their own racial or ethnic background was calculated at 3.02 with a standard deviation of 0.85. The results for statement 16 are displayed in Table 7.

Table 7

Student Survey Statement 16 Responses

Student Demographics Category	Student Response	*f*	%
All Students	Strongly Agree or Agree	47	19.19
	Neutral	126	51.43
	Disagree or Strongly Disagree	72	29.39

Note. $N = 245$, f = frequency, % = percentage.

Faculty demographic statements. Faculty respondents were asked to provide their ages by selecting one of four options: 25-35, 36-50, 51-64, or 65 and older. The respondents identified themselves as follows: 25-35 (17.65%), 36-50 (35.29%), 51-64 (32.35%), and 65 and over (14.71%). Most respondents were females (58.82%) compared to males (41.18%). The final demographic question was about race, and respondents selected one of the following categories: White, Black/African American, Hispanic/ Latino, Asian/Pacific Islander, Native American/American Indian, or Other. Most respondents were White (82.35%) and Hispanic (11.76%). Only 2.94% of the respondents selected Asian/Pacific Islanders or Native American/American Indian, and there were no Black/African American participants. Faculty respondent demographics are displayed in Table 8.

Table 8

Faculty Respondent Demographics

Demographic Category	Option	*f*	%
Age	25-35	6	17.65
	36-50	12	35.29
	51-64	11	32.35
	65+	5	14.71
Gender	Female	20	58.82
	Male	14	41.18
Race	White	28	82.35
	Black/African American	0	0
	Hispanic/Latino	4	11.76
	Asian/Pacific Islander	1	2.94
	Native American/American Indian	1	2.94
	Other	0	0

Note. $N = 34$, f = frequency, % = percentage.

Faculty survey statement one. *The course evaluation helps improve my teaching effectiveness.* For survey statement one, 79.41% of respondents strongly agreed or agreed, 17.65% selected neutral, and 2.94% indicated disagree or strongly disagree. Male respondents and female respondents had similar responses, with 80% of males indicating strongly agree or agree and 78.57% of females indicating strongly agree or agree. Another 15% of females and 21.43% of males indicated neutral. No male respondents indicated disagree or strongly disagree, and only 5% of female respondents indicated disagree or strongly disagree.

When disaggregating the data by age range, respondents selected strongly agree or agree as follows: 25-35 (66.67%), 36-50 (75%), 51-64 (90.91%), and 65 and older (80%). Neutral was indicated by the following: 25-35 (33.33%), 36-50 (16.67%), 51-64 (.09%), and 65 and older (20%). Disagree and strongly disagree were only indicated by those respondents age 36-50 (8.33%).

When disaggregating the data by race, most respondents selected strongly agree or agree (White = 78.57% and Non-White = 83.33%). Neutral was selected as a response by 17.86% of White participants and 16.67% of Non-White participants. Disagree or strongly disagree was only selected by White respondents (3.57%); no respondents identifying as Non-White selected disagree or strongly disagree as their response to this statement. Based on the results of all respondents, a mean score with regard to the SET improving teaching effectiveness was calculated at 2.18 with a standard deviation of 0.71.

Faculty survey statement two. *The course evaluation form is adequate to evaluate my teaching effectiveness.* For survey statement two, 8.82% of respondents strongly agreed, 61.76% agreed, 23.53% selected neutral, and only 5.88% disagreed.

Male and female respondents had similar responses, with 71.43% of males indicating strongly agree or agree and 70% of females indicating strongly agree or agree. Another 25% of females and 21.43% of males indicated neutral. Only 7.14% of male respondents indicated disagree or strongly disagree compared to 5% of the female respondents.

When disaggregating the data by age range, respondents selected strongly agree or agree as follows: 25-35 (66.67%), 36-50 (66.66%), 51-64 (63.64%), and 65 and older (100%). Neutral was indicated by the following: 25-35 (33.33%), 36-50 (25%), and 51-64 (27.27%). No respondents age 65 or older selected neutral. Disagree and strongly disagree were only indicated as the responses to this statement by respondents age 36-50 (8.33%) and 51-64 (9.09%).

When disaggregating the data by race, most respondents selected strongly agree or agree (White = 71.43% and Non-White = 66.67%). Neutral was selected as a response by 21.43% of White participants and 33.33% of Non-White participants. Disagree or strongly disagree was only selected by respondents identifying as White (7.14%). Based on the results of all respondents, a mean score with regard to the SET's ability to evaluate professor performance was calculated at 2.26 with a standard deviation of .070.

Faculty survey statement three. *Students should take student evaluations of teaching seriously.* Over 90% of respondents agreed the SET should be taken seriously by students. For this survey item, 50% strongly agreed, 44.12% agreed, and 5.88% selected neutral. No respondents indicated they disagreed. All male respondents (100%) selected strongly agree or agree, while 90% of female respondents (90%) agreed or strongly agreed. There was no difference when disaggregating by age, with more than 90% of all respondents selecting strongly agree or disagree. Only 16.67% of respondents

age 36-50 selected neutral, and no other age group selected neutral as a response to this statement.

When disaggregating the data by race, most respondents selected strongly agree or agree (White = 96.43% and Non-White = 83.33%). Neutral was selected as a response by 3.57% of White participants and 16.67% of Non-White participants. Based on the results of all respondents, a mean score with regard to student preference for male instructors was calculated at 1.56, with a standard deviation of 0.60.

Faculty survey statement four. *Students read and understand each statement before they rate it.* Nearly half (47.06%) of faculty respondents believed students read and understand each statement. For this survey item, 20.59% strongly agreed, 26.47% agreed, 35.29% selected neutral, 14.71% disagreed, and 2.94% strongly disagreed. Male respondents and female respondents had similar responses, with 50% of males and 45% of females indicating strongly agree or agree. However, there was a difference for respondents who indicated being neutral (males = 21.43% and females = 45%). A significant difference was also found between the males (28.57%) and females (10%) who indicated disagree or strongly disagree.

When disaggregating the data by age, respondents selected strongly agree or agree as follows: 25-35 (66.66%), 26-40 (33.33), 41-55 (63.63%), and 56 or over (20%). Neutral was indicated as the response by the following age groups: 25-35 (33.33%), 26-40 (41.67), 41-55(18.18%), and 56 or over (60%). When disaggregating the data by race, respondents indicated strongly agree or agree as follows: White (50%) and Non-White (33.33%). Neutral was selected by White (32.14%) and Non-White (50%) participants. Disagree and strongly disagree were indicated as the responses by 17.85% of White participants and 16.67% of Non-White participants. Based on the results of all

respondents, a mean score with regard to faculty perception of students' understanding of the SET statements was calculated at 2.53 with a standard deviation of 1.06.

Faculty survey statement five. *Students prefer taking courses from male instructors.* For survey statement five, respondents selected agree (2.94%), neutral (64.71%), disagree (26.47%), and strongly disagree (5.88%). When disaggregating the data by gender, most respondents selected neutral as a response to this statement (males = 71.43% and females = 60%). Disagree or strongly disagree was selected by 28.57% of males and 35% of females. No male respondents selected strongly agree or agree, and only 5% of the females selected agree as their response to this statement.

When disaggregating the data by age, agree was only selected by age group 25-35 (16.67%). Most respondents selected neutral in response to this statement as follows: 25-35 (50%), 36-50 (66.67%), 51-64 (54.55%), and 65 and over (100%). Respondents selected disagree or strongly disagree with the following frequencies: 25-35 (33.34%), 36-50 (33.33%), and 51-64 (45.45%). No respondents age 65 or older selected disagree or strongly disagree as a response to this statement.

When disaggregating the data by race, only White respondents (3.57%) indicated strongly agree or agree as a response to this statement. Neutral was selected as the response by both White (71.43%) and Non-White (33.33%) participants. Disagree and strongly disagree were indicated by 25% of White participants and 66.67% of Non-White participants. Based on the results of all respondents, a mean score with regard to student preference for male instructors was calculated at 3.35 with a standard deviation of 0.64. The results for faculty statement five responses by gender are displayed in Table 9.

Table 9

Summary of Faculty Statement Five Responses by Gender

Gender	Faculty Response	*f*	%
Male	Strongly Agree or Agree	0	0
	Neutral	10	71.43
	Disagree or Strongly Disagree	4	28.57
Female	Strongly Agree or Agree	1	5
	Neutral	12	60
	Disagree or Strongly Disagree	20	35

Note. $N = 34$, *f* = frequency, % = percentage.

Faculty survey statement six. *Students prefer taking courses from female instructors.* Similar to survey statement five, most faculty respondents (67.65%) selected neutral, 26.47% selected disagree, 5.88% selected strongly disagree, and no respondents selected strongly agree or agree. When disaggregating the data by gender, most respondents selected neutral (males = 71.43% and females = 65%). Disagree and strongly disagree were selected by 28.57% of males and 35% of females.

When disaggregating the data by age, most respondents selected neutral with the following percentages: 25-35 (66.67%), 36-50 (66.67%), 51-64 (54.55%), and 65 and over (100%). Disagree or strongly disagree was selected as follows: 25-35 (33.34%), 36-50 (33.33%), and 51-64 (45.45%). No respondents 65 or over selected disagree or strongly disagree.

When disaggregating the data by race, most White respondents (75%) selected neutral compared to Non-Whites (33.33%). Disagree or strongly disagree was selected as a response to this statement by 66.67% of Non-White participants and 25% of White participants. Based on the results of all respondents, a mean score with regard to student

preference for female instructors was calculated at 3.38 with a standard deviation of 0.59. The results for faculty statement six responses by gender are displayed in Table 10.

Table 10

Summary of Faculty Statement Six Responses by Gender

Gender	Faculty Response	*f*	%
Male	Strongly Agree or Agree	0	0
	Neutral	10	71.43
	Disagree or Strongly Disagree	3	28.57
Female	Strongly Agree or Agree	0	0
	Neutral	13	65
	Disagree or Strongly Disagree	7	35

Note. $N = 33$, *f* = frequency, % = percentage.

Faculty survey statement seven. *Students prefer taking courses from young and enthusiastic instructors.* For survey statement seven, 2.94% of respondents strongly agreed, 44.12% agreed, 44.12% selected neutral, 8.82% disagreed, and no respondents selected strongly disagree as a response to this statement. When disaggregating the data by gender, 50% of female respondents selected strongly agree or agree compared to 42.86% of male respondents. Neutral was selected as a response as follows: female (50%) and male (35.77%). Only male respondents selected disagree (21.43%).

When disaggregating the data by age, respondents selected strongly agree or agree with the following percentages: 25-35 (50%), 36-50 (41.67%), 51-64 (54.55%), and 65 and over (40%). Neutral was the response as follows: 25-35 (33.33%), 36-50 (50%), 51-64 (36.36%), and 65 or older (60%). Disagree and strongly disagree was indicated by 16.67% of those age 25-35, 8.33% of those age 36-50, and 9.09% of those age 51-64. When disaggregating the data by race, respondents indicated strongly agree or agree as follows: White (46.43%) and Non-White (50%). Neutral was selected by 46.43% of

Whites and 33.33% of Non-Whites. Disagree or strongly disagree was indicated by 7.14% of White participants and 16.67% of Non-White participants. Based on the results of all respondents, a mean score with regard to student preference for young and enthusiastic instructors was calculated at 2.59 with a standard deviation of 0.69.

Faculty survey statement eight. *Students prefer taking courses from older and experienced instructors.* For this survey statement, 2.94% strongly agreed, 41.18% agreed, 38.24% selected neutral, and 17.65% disagreed. No respondents strongly disagreed. When disaggregating the data by gender, both female respondents (45%) and male respondents (42.86%) selected strongly agree and agree. Neutral was selected as the response to this statement by 45% of females and 28.57% of males. Disagree or strongly disagree was selected as follows: females (10%) and males (28.57%).

When disaggregating the data by age, respondents selected strongly agree or agree with the following percentages: 25-35 (33.33%), 36-50 (25%), 51-64 (72.73%), and 65 or over (40%). Neutral was the response as follows: 25-35 (50%), 36-50 (50%), 51-64 (18.18%), and 65 or over (40%). Respondents selected disagree or strongly disagree with the following frequencies: 25-35 (16.67%), 36-50 (25%), 51-64 (9.09%), and 65 and older (20%).

When disaggregating the data by race, respondents indicated strongly agree or agree as a response to this statement as follows: White (42.86%) and Non-White (50%). Neutral was selected by 42.86% of Whites and 16.67% of Non-Whites. Disagree or strongly disagree was indicated by 14.29% of White participants and 33.33% of Non-White participants. Based on the results of all respondents, a mean score with regard to student preference for older and experienced instructors was calculated at 2.71 with a standard deviation of 0.79.

Faculty survey statement nine. *Students are comfortable taking courses from instructors who speak English as their second language.* For this survey statement, 2.94% strongly agreed, 23.53% agreed, 38.24% selected neutral, 29.41% disagreed, and 5.88% strongly disagreed. When disaggregating the data by gender, both female respondents (20%) and male respondents (35.71%) selected strongly agree or agree. Neutral was selected as follows: females (45%) and males (28.57%). Disagree or strongly disagree was indicated by 35% of females and 35.71% of males. When disaggregating the data by age, respondents selected strongly agree or agree with the following percentages: 25-35 (33.33%), 36-50 (25%), 51-64 (27.27%), and 65 or over (20%). Neutral was the response as follows: 25-35 (33.33%), 36-50 (50%), 51-64 (27.27%), and 65 and over (40%). Respondents selected disagree or strongly disagree with the following frequencies: 25-35 (33.33%), 36-50 (25.50%), 51-64 (45.45%), and 65 or older (40%).

When disaggregating the data by race, respondents indicated strongly agree or agree as follows: White (25%) and Non-White (33.33%). Neutral was selected by 35.71% of Whites and 50% of Non-Whites. Disagree or strongly disagree was indicated by the following: White (39.28%) and Non-White (16.67%). Based on the results of all respondents, a mean score with regard to student comfort with instructors who speak English as their second language was calculated at 3.12 with a standard deviation of 0.93. The results for faculty statement nine responses by race are displayed in Table 11.

Table 11

Summary of Faculty Statement Nine Responses by Race

Faculty Race	Faculty Response	*f*	%
White	Strongly Agree or Agree	16	57.15
	Neutral	4	14.29
	Disagree or Strongly Disagree	8	28.57
Non-White	Strongly Agree or Agree	3	50
	Neutral	1	16.67
	Disagree or Strongly Disagree	2	33.33

Note. $N = 33$, *f* = frequency, % = percentage. Non-White = Hispanic/Latino, Asian/Pacific Islander, Native American/American Indian, Black/African American, and/or Other.

Faculty survey statement 10. *Students are more comfortable taking courses from instructors who speak English as their first language.* More than half of faculty participants (67.65%) responded students are more comfortable taking courses from instructors who speak English as their first language. For this survey item, 20.59% of respondents strongly agreed, 47.06% agreed, 29.41% selected neutral, 2.94% disagreed, and no respondents strongly disagreed with the statement. Male respondents and female respondents had similar responses, with 71.43% of males indicating strongly agree or agree and 65% of females indicating strongly agree or agree. Another 35% of the females and 21.43% of the males selected neutral. Only 7.14% of male respondents indicated disagree or strongly disagree, and no female respondents indicated disagree.

When disaggregating the data by age, respondents selected strongly agree or agree as follows: 25-35 (66.67%), 36-50 (58.33%), 51-64 (81.82%), and 65 or over (60%). Neutral was selected with the following percentages: 25-35 (16.67%), 36-50 (41.67%), 51-64 (18.18%), and 65 and over (40%). Disagree or strongly disagree was selected as a

response by 16.67% of those age 25-35. No other respondents selected disagree or strongly disagree as a response.

When disaggregating the data by race, respondents indicated strongly agree or agree as follows: White (71.43%) and Non-White (50%). Neutral was selected by 25% Whites and 50% by Non-Whites. Only White respondents (3.57%) indicated disagree or strongly disagree as a response to this statement. Based on the results of all respondents, a mean score with regard to student preference for instructors who speak English as their first language was calculated at 2.15 with a standard deviation of 0.77.

Faculty survey statement 11. *Students pay more attention to instructor personality (i.e., friendliness, leniency, looks, or dress) than teaching effectiveness.* For survey statement 11, over half of faculty respondents (55.89%) reported students pay significant attention to their instructor's personality. Respondents selected strongly agree (14.71%), agree (41.18%), neutral (14.71%), disagree (23.53%), and strongly disagree (5.88%). Male and female respondents had similar responses, with 57.14% of the males indicating strongly agree or agree compared to 55% of the females. Neutral was selected by 14.29% of males and 15% of females. Disagree or strongly disagree was selected as follows: females (30%) and males (28.57%).

When disaggregating the data by age, respondents selected strongly agree or agree with the following frequencies: 25-35 (50%), 36-50 (50%), 51-64 (54.54%), and 65 or over (80%). Neutral was selected as follows: 25-35 (16.67%), 36-50 (33.33%), and no respondents 51 or older selected neutral as their response. Respondents selected disagree or strongly disagree with the following percentages: 25-35 (33.33%), 36-50 (16.67%), 51-64 (45.45%), and 65 or older (20%).

When disaggregating the data by race, respondents indicated strongly agree or agree as follows: White (57.15%) and Non-White (50%). Neutral was selected by 14.29% of Whites and 16.67% of Non-Whites. Disagree or strongly disagree was indicated as a response by 28.57% of White participants and 33.33% of Non-White participants. Based on the results of all respondents, a mean score with regard to student focus on an instructor's personality (i.e., friendliness, leniency, looks, or dress) was calculated at 2.65 with a standard deviation of 1.16. The results for faculty statement 11 responses are displayed in Table 12.

Table 12

Faculty Survey Statement 11 Responses

Faculty Demographics Category	Faculty Responses	*f*	%
All Faculty	Strongly Agree or Agree	19	55.89
	Neutral	5	14.71
	Disagree or Strongly Disagree	10	29.41

Note. $N = 34$, *f* = frequency, % = percentage.

Faculty survey statement 12. *Students prefer instructors who are assertive and provide clear expectations for the course.* For survey statement 12, the most-frequent response was agree (67.65%), followed by strongly agree (26.47%), neutral (2.94%), and disagree (2.94%). No respondents selected strongly disagree. Male and female respondents had similar responses, with 92.86% of males and 95% of females indicating strongly agree or agree. No male respondents selected neutral, and only 5% of females selected neutral as a response to this statement. No female respondents indicated disagree or strongly disagree, while 7.14% of male respondents indicated disagree or strongly disagree.

When disaggregating the data by age range, all respondents selected strongly agree or agree except those in the 36-50 age range who selected neutral (8.33%) and

disagree (8.33%). When disaggregating the data by race, respondents selected strongly agree or agree as follows: White (96.43%) and Non-White (83.33%). Neutral was only selected as a response to this statement by Non-White respondents (16.67%). Disagree or strongly disagree was only selected by White respondents (3.57%). Based on the results of all respondents, a mean score with regard to student preference for assertive instructors was calculated at 1.82 with a standard deviation of 0.62.

Faculty survey statement 13. *Students prefer instructors who are caring and concerned about them outside of the classroom.* For survey question 13, over two-thirds (76.47%) of the respondents reported students prefer instructors who are caring and concerned about them outside of the classroom. Respondents selected strongly agree (29.41%), agree (47.06%), neutral (20.59%), and disagree (2.94%). No respondents selected strongly disagree. Male and female respondents had similar responses, with 71.43% of males and 80% of females indicating strongly agree or agree. Neutral was indicated by 21.43% of males and 20% of females. Disagree was only selected as a response to this statement by males (7.14%).

When disaggregating the data by age, respondents selected strongly agree or agree as follows: 25-35 (100%), 36-50 (66.67%), 51-64 (72.72%), and 65 or over (80%). Neutral was selected with the following percentages: 36-50 (25%), 51-64 (27.27%), and 65 or older (20%). No respondents age 25-35 selected neutral. Disagree or strongly disagree was only indicated as a response to this statement by those age 36-50 (8.33%). When disaggregating the data by race, respondents indicated strongly agree or agree with the following frequencies: White (75%) and Non-White (83.33%). Neutral was only selected by White respondents (25%). Disagree or strongly disagree was indicated as the response by Non-Whites (33.33%). Based on the results of all respondents, a mean score

in regard to student preference for caring instructors was calculated at 1.97 with a standard deviation of 0.79.

Faculty survey statement 14. *Many students have taken a class from an instructor of a different race or ethnicity.* For survey statement 14, 61.29% of respondents agreed students had taken a class from an instructor of a different race or ethnicity compared to 38.71% of respondents who indicated disagree as a response to this statement (see Table 13). When disaggregating the data by gender, male respondents (57.14%) and female respondents (64.71%) indicated they believed many students had taken a class from an instructor of a different race or ethnicity.

When disaggregating the data by age, respondents selected agree with the following percentages: 25-35 (80%), 36-50 (70%), 51-64 (54.55%), and 65 or over (40%). Respondents selected disagree as follows: 25-35 (20%), 36-50 (30%), 51-64 (45.45%), and 65 or older (60%). When disaggregating the data by race, respondents indicated agree with the following frequencies: White (56%) and Non-White (83.33%). Disagree was indicated by 44% of Whites and 16.67% of Non-Whites. Based on the results of all respondents, a mean score with regard to student exposure to an instructor from a different race or ethnicity was calculated at 1.39 with a standard deviation of 0.49.

Table 13

Faculty Survey Statement 14 Responses

Faculty Demographics Category	Faculty Response	*f*	%
All Faculty	Agree	19	61.29
	Disagree	12	38.71

Note. $N = 34$, *f* = frequency, % = percentage.

Faculty survey statement 15. *Many students are excited to take a class from an instructor from a different race or ethnicity than their own.* The most-frequent response

to this statement was neutral (73.53%), followed by agree (17.65%), strongly agree (5.88%), and disagree (2.94%). No respondents selected strongly disagree as a response to this statement. When disaggregating the data by gender, both female respondents (25%) and male respondents (21.43%) selected strongly agree or agree. Neutral was selected as a response by 75% of females and 71.43% of males. Disagree or strongly disagree was only selected by male respondents (7.14%).

When disaggregating the data by age, respondents selected strongly agree or agree as follows: 25-35 (33.34%), 36-50 (91.66%), 51-64 (18.18%), and 65 and over (60%). Those responding neutral to this statement included the following: 25-35 (66.67%), 36-50 (8.33%), 51-64 (81.82%), and 65 and over (40%). No respondents selected disagree or strongly disagree as a response to this statement. When disaggregating the data by race, respondents indicated strongly agree or agree with the following percentages: White (25%) and Non-White (16.67%). Neutral was selected by 71.43% of Whites and 83.33% of Non-Whites. Disagree or strongly disagree was only indicated as a response to this statement by White respondents (3.57%). Based on the results of all respondents, a mean score with regard to student preference for male instructors was calculated at 2.74 with a standard deviation 0.61.

Faculty survey statement 16. *Many students are more comfortable with instructors who appear to be from their own racial or ethnic background.* For this survey statement, 14.71% strongly agreed, 38.24% agreed, 41.18% selected neutral, 5.88% disagreed, and no respondents selected strongly disagree. When disaggregating the data by gender, female respondents (55%) and male respondents (50%) selected strongly agree or agree. Neutral was selected as follows: females (40%) and males (42.86%). Disagree or strongly disagree was selected by 5% of females and 7.14% of males.

When disaggregating the data by age, respondents selected strongly agree or agree as follows: 25-35 (33.33%), 36-50 (33.34%), 51-64 (72.72%), and 65 and over (80%). Neutral was the response with the following percentages: 25-35 (33.33%), 36-50 (66.67%), 51-64 (27.27%), and 65 and over (20%). Only respondents age 25-35 (33.33%) selected disagree or strongly disagree as a response to this statement.

When disaggregating the data by race, respondents indicated strongly agree or agree with the following frequencies: White (57.15%) and Non-White (16.67%). Neutral was selected by 35.71% of Whites and 66.67% of Non-Whites. Disagree or strongly disagree was only indicated as a response to this statement by White respondents (7.14%). Based on the results of all respondents, a mean score with regard to student preference for instructors who appear to be from their own racial or ethnic background was calculated at 2.38 with a standard deviation 0.80. The results for faculty statement 16 responses are displayed in Table 14.

Table 14

Faculty Survey Statement 16 Responses

Faculty Demographics Category	Faculty Responses	*f*	%
All Faculty	Strongly Agree or Agree	18	52.95
	Neutral	14	41.18
	Disagree or Strongly Disagree	2	5.88

Note. $N = 34$, *f* = frequency, % = percentage.

Summary

The purpose of this quantitative study was to determine the degree of influence of implicit bias on the SET. In this chapter, the student and faculty survey results were analyzed and presented. In Chapter Five, the findings are presented, and the three research questions are discussed within the conclusions section. Implications for practice are presented, and recommendations for further research are proposed.

Chapter Five: Summary and Conclusions

The intention of this study was to determine the impact of implicit bias on the SET at a Missouri community college. The SET has been the standard measurement of teacher effectiveness since the 1920s (Degheri, 2017; Wachtel, 1998). Faculty and administrators have questioned the validity and reliability of the SET results since its inception (Spooren & Christiaens, 2017). Bias in the SET results has been a prevalent concern over the last 20 years as the numbers of marginalized faculty have increased (Prasad et al., 2017). Women and people of color have voiced their concerns about student comments regarding race, ethnicity, and physical appearance allowed and considered as part of the ratings (Prasad et al., 2017).

This quantitative study was conducted by surveying faculty in the Communication and World Languages department at a Missouri community college and the students enrolled in their classes during the fall 2019 semester. A review of literature provided evidence of a body of research about the influence of implicit bias on the SET. However, few researchers have examined the impact of the SET regarding race and non-native English speakers (Wallace et al., 2019). With the increase in diverse faculty members and students on college campuses across the country, it is important to recognize and reduce implicit bias to promote a more culturally sensitive and tolerant society (Smith & Hawkins, 2011). Results from the SET have been used for personnel decisions, including promotion, tenure, and termination, which have created potential issues for diverse faculty members (Mitchell & Martin, 2018).

Colleges and universities have used the SET to demonstrate accountability of instructor quality; however, the SET may actually measure student satisfaction, not teaching effectiveness (Spooren & Christiaens, 2017). Furthermore, there has been no

consensus among educators defining teaching effectiveness (Sauer, 2012). While these factors present concern about the SET, the SET remains a common tool for faculty performance assessment (Hornstein, 2017).

In Chapter Five, the findings of the quantitative data collected to analyze the influence of implicit bias on the SET are presented. Following the findings, conclusions based on these findings are discussed. Finally, implications for practice are described, and recommendations for future research are suggested.

Findings

The findings of the study serve as a foundation for a broader understanding of implicit bias on the SET. The data from this study were collected from a Likert-type survey administered to faculty, and a similar survey was administered to the students enrolled in their classes. In the following section, the results from the analysis of data are presented.

Research question one. *What percentage of students utilize a different criterion to evaluate faculty from marginalized groups than faculty from non-marginalized groups?* This research question was answered based upon data from student survey statements six, eight, nine, and 15. The following is a summary of the results.

Student survey statement six. *I prefer taking courses from female instructors.* Most students selected neutral, indicating no preference or objection to female instructors. Male students indicated a preference for female instructors at a greater percentage (17.73%) than female students (12.72%). Fewer than 9% of students disagreed with the statement.

Student survey statement eight. *I prefer taking courses from older and experienced instructors.* Approximately half of the student respondents selected neutral,

indicating no preference or objection for older and experienced instructors. Interestingly, students in the age range of 18-25 (36.73%) and in the age range of 26-40 (43.34%) indicated a preference for older instructors. Only 20% of students age 41 and older indicated a preference for older and experienced instructors.

Student survey statement nine. *I am comfortable taking courses from instructors who speak English as their second language.* Overall, half of the respondents (50.2%) agreed with this statement. However, more than 20% of respondents indicated they were uncomfortable taking courses from instructors who speak English as their second language. Students identifying as Non-White, which included Black/African American, Hispanic/Latino, Asian/Pacific Islander, Native American, or Other (50.2%) agreed to this statement more than students identifying as White (47.39%).

Student survey statement 15. *I am excited to take a class from an instructor from a different race or ethnicity than my own.* Approximately half (51.83%) of the overall respondents agreed with this statement. Similar to the results from survey statement nine, students identifying as Non-White indicated greater interest in taking a class from an instructor of a different racial or ethnic background (73.53%) compared to students identifying as White (48.34%).

Research question two. *Based on the opinions of students and faculty, to what extent does implicit bias influence the outcome of student evaluations of teaching (SET)?* This research question was answered based upon data from student and faculty survey statements five, seven, 10, and 16. The following is a summary of the results.

Student survey statement five. *I prefer taking courses from male instructors.* Most students selected neutral, indicating no preference for or objection to male instructors. Of those students indicating a preference, females surpassed males by

selecting agree at a greater percentage (11.59%) than males (8.86%). Overall, 12.71% of students disagreed with the statement.

Student survey statement seven. *I prefer taking courses from young and enthusiastic instructors.* Many students selected neutral as a response to this statement, indicating no preference for young and enthusiastic instructors. Traditional college age students (18-25) indicated a preference for younger instructors more than the other age groups (39.49%).

Student survey statement 10. *I am more comfortable taking courses from instructors who speak English as their first language.* Over 59% of students indicated being more comfortable taking courses from instructors who speak English as their first language. Few students (4.89%) indicated they disagreed with this statement. Interestingly, more students who identified as White selected agree to this statement (61.13%) than students identifying as Black/African American, Hispanic/Latino, Asian/Pacific Islander, Native American/American Indian, or Other (47.06%).

Student survey statement 16. *I am more comfortable with instructors who appear to be from my own racial or ethnic background.* Overall, half of the students (51.43%) selected neutral as the response to this statement. Students identifying as Black/African American selected agree more than any other group (80%), followed by students identifying as Hispanic/Latino (33.33%).

Faculty survey statement five. *Students prefer taking courses from male instructors.* Most faculty respondents (64.71%) selected neutral as their response to this statement. Interestingly, no male respondents agreed with this statement, and nearly 30% of male respondents selected disagree as their response. Only 5% of female respondents agreed with this statement.

Faculty survey statement seven. *Students prefer taking courses from young and enthusiastic instructors.* Nearly half of the faculty respondents (47.06%) agreed with this statement. In comparison, the other half of the faculty respondents (44.12%) selected neutral as their response. Of those faculty respondents selecting disagree (16.67%), most were in the youngest age group (25-35).

Faculty survey statement 10. *Students are more comfortable taking courses from instructors who speak English as their first language.* More than half of faculty respondents (67.65%) believed students are more comfortable taking courses from instructors who speak English as their first language. Interestingly, more faculty identifying as White (71.43%) selected agree than did Non-White faculty respondents (50%).

Faculty survey statement 16. *Many students are more comfortable with instructors who appear to be from their own racial or ethnic background.* Overall, half of the faculty respondents (52.95%) selected agree as a response to this statement. Faculty respondents identifying as White selected agree at a greater percentage (57.15%) than faculty respondents identifying as Non-White (33.34%).

Research question three. *Based on the opinions of students and faculty, to what extent do instructor characteristics predict the outcome of student evaluations of teaching (SET)?* This research question was answered based upon data from survey statements 11, 12, and 13 presented to students and faculty. The following is a summary of the results.

Student survey statement 11. *When evaluating professors, I pay more attention to their personality (i.e., friendliness, leniency, looks, or dress).* More than half of student respondents (50.61%) selected agree as a response to this statement. Males selected agree at a slightly higher percentage (51.90%) than females (49.70%).

Traditional college-age students (18-25) selected agree to this statement at a higher percentage (54.60%) than any other age group.

Student survey statement 12. *I prefer instructors who are assertive and provide clear expectations for the course.* Overall, students selected agree (85.65%) more than the other options for this statement. Only 11.89% of respondents indicated neutral, and 2.45% selected disagree.

Student survey statement 13. *I prefer instructors who are caring and concerned about me outside of the classroom.* Similar to survey statement 12, students indicated agree at more than 80%. Only 14.69% of respondents indicated neutral, and 2.45% selected disagree.

Faculty survey statement 11. *Students pay more attention to instructor personality (i.e., friendliness, leniency, looks, or dress) than teaching effectiveness.* More than half of faculty respondents (55.89%) believed students pay more attention to instructor characteristics than teaching effectiveness. Males selected agree at a slightly higher percentage (57.14%) than females (55%).

Faculty survey statement 12. *Students prefer instructors who are assertive and provide clear expectations for the course.* Overwhelming, faculty respondents (94.12%) reported the belief students prefer assertive instructors who provide clear expectations. Interestingly, only male respondents selected disagree as a response to this statement.

Faculty survey statement 13. *Students prefer instructors who are caring and concerned about them outside of the classroom.* Most faculty respondents (76.47%) selected agree as their response to this statement. Female respondents selected agree at a higher percentage (80%) than males (71.43%). Only male faculty respondents disagreed with this statement (7.14%).

Conclusions

The intent of this study was to determine the extent to which implicit bias influences the outcome of the SET. This study was grounded in Greenwald and Banaji's (1995) implicit bias theory that people unconsciously perceive those who are different in a negative way (Staats et al., 2015). The impact of implicit bias is a national conversation not only in the field of education but also in health care, criminal justice, employment, housing, and other areas across the country (Staats et al., 2017).

In the following paragraphs, each research question is stated. The results of the faculty and student surveys related to each research question are presented. Finally, the results are compared to the current literature surrounding implicit bias.

Research question one. *What percentage of students utilize a different criterion to evaluate faculty from marginalized groups than faculty from non-marginalized groups?* The results from the survey did not support the premise that students use a different criterion to evaluate faculty based on gender. Student respondents selected neutral as their response to survey statements regarding preference for male or female instructors at nearly 80%. In addition, more than 10% stated they had no preference for either gender.

These specific findings did not support the literature presented in Chapter Two. For example, Wallace et al. (2019) found female instructors were rated lower than men in the areas of competence, organization, and professionalism. Rosen (2017) concluded gender bias is a serious concern for women because men are rated higher on the SET in every discipline. Boring et al. (2016) suggested male students give male instructors higher ratings than female instructors. While the current study was noted based upon

actual SET results, the survey results provided a reason to conclude instructor gender may not influence SET responses.

The results from the survey did not support the premise that students use a different criterion to evaluate faculty based on age. Student respondents selected neutral as their response to survey statements regarding the preference for younger or older instructors at nearly 60%. In addition, fewer than 5% of students indicated they were opposed to older and experienced instructors. More than 30% of student respondents indicated a preference for older and more experienced instructors as opposed to younger and enthusiastic instructors.

These findings were inconsistent with the literature presented in Chapter Two. For example, Doubleday and Lee (2016) found students gave lower ratings to older instructors than younger instructors for a narrated video containing identical content. Arbuckle and Williams (2003) discovered similar results when they had students rate a recorded lecture containing the same content. The only difference was the age and gender of the instructor.

The survey results supported the premise students use a different criterion to evaluate faculty based on English proficiency. More than 20% of respondents indicated they were uncomfortable taking courses from instructors who speak English as their second language. Early studies by Kulik and Kulik (1974) revealed highly-rated instructors were characterized as verbally fluent and exhibited effective communication skills. In a recent study by Fan et al. (2019), results indicated a bias against instructors who speak English as their second language. Both studies supported the probability of implicit bias impacting results of the SET.

The survey results were inconclusive regarding student use of a different criterion to evaluate faculty based on race and ethnicity. Approximately 50% of students indicated neutral as a response to statements affirming the desire to take a course from an instructor of a different race or ethnicity, while the other half wanted to take a course from an instructor of a different race or ethnicity. Fewer than 3% of the students indicated they did not want to take a course from an instructor of a different race or ethnicity. While the research presented in Chapter Two indicated quantitative data on race and the SET are limited and inconclusive (Huston, 2006; Reid, 2010), this current study demonstrated some students are interested in experiencing instructors from diverse backgrounds.

Research question two. *Based on the opinions of students and faculty, to what extent does implicit bias influence the outcome of student evaluations of teaching (SET)?* Student opinion indicated gender does not significantly influence responses on the SET; therefore, implicit bias has a minimal impact on the outcome of the SET. As stated in the findings to research question one, most students indicated no preference or objection to male or female instructors. Faculty opinion also indicated gender does not significantly influence responses on the SET. Only 3% of faculty members believed students prefer male instructors.

These findings were inconsistent with the literature presented in Chapter Two. For instance, McPherson et al. (2009) found bias for male, younger, and white instructors on the SET. Bavishi et al. (2010) asserted women are marginalized based on gender role expectations and stereotypes.

Student opinion indicated age may influence responses on the SET. Nearly 35% of students indicated a preference for both older and younger instructors; however, many students (60%) indicated no preference for older or younger instructors. Faculty opinion

also indicated age might impact the outcome of the SET. Approximately 50% of faculty participants believed students prefer older faculty, while 50% of faculty participants believed students prefer younger instructors.

The literature in Chapter Two indicated age is a factor in SET outcomes. For example, McPherson and Jewell (2007) found SET scores for older instructors decreased even though students indicated the need for more experienced instructors. The current study did not provide sufficient evidence to conclude that age does or does not have a significant impact on the SET.

Student opinion indicated English proficiency influences responses on the SET; therefore, based on the students' responses, implicit bias has an impact on the outcome of the SET. Nearly 60% of the students indicated a preference for instructors who speak English as their first language. Faculty opinion also indicated English proficiency influences responses on the SET. Interestingly, faculty believed students are more comfortable with language differences than the students indicated on the student survey. These findings were consistent with the literature presented in Chapter Two. Huston (2006) found non-native English speakers are rated lower on the SET than native speakers.

Fan et al. (2019) revealed a significant bias against faculty with non-English speaking backgrounds. Student opinion indicated race and ethnicity did not significantly influence responses on the SET. Fewer than 20% of student respondents indicated they felt more comfortable with instructors who appeared to be from their own racial and ethnic background. In addition, nearly 30% rejected the idea of being more comfortable with instructors from the same racial and ethnic background. Student respondents were also asked if they have had a class with an instructor from a different racial or ethnic

background. Over 40% of student respondents indicated they had never had an instructor from a different racial or ethnic background; therefore, the ability to judge their comfort level with an instructor from a different race or ethnicity cannot be determined.

Faculty opinion was very different from student opinion, with over 50% of faculty respondents indicating they believed students were more comfortable with instructors from their own racial and ethnic backgrounds. Nearly 60% of the faculty identifying as White believed students were more comfortable with their own racial and ethnic background, while only 33% of those identifying as Non-White believed this to be true. In addition, fewer than 6% of faculty rejected the idea of students being more comfortable with faculty from their own racial and ethnic backgrounds.

As stated in Chapter Two, studies about race have been limited and inconclusive. Basow et al. (2013) suggested research limitations are due to the limited number of non-white faculty. Reid (2010) concluded the lack of student exposure to faculty of color, particularly faculty identifying as Black, contributes to bias and stereotypical comments on the SET. In the current study, only 17.65% of faculty identified as Non-White, and no faculty participants identified as Black/African American. The lack of representation from faculty of color does create an inability to offer a greater perspective on this issue.

Research question three. *Based on the opinions of students and faculty, to what extent do instructor characteristics predict the outcome of student evaluations of teaching (SET)?* Instructor characteristics, including personality, style of dress, leniency, and physical appearance influenced how students rate instructors. More than 50% of student respondents indicated those traits influence their responses on the SET, more than the qualities of teaching effectiveness. In addition, more than 55% of faculty respondents reported students rate instructors based on characteristics, not on teaching effectiveness.

These findings were consistent with the literature reviewed in Chapter Two. Early researchers Stalnaker and Remmers (1928) acknowledged the Purdue Rating Scale is subject to the halo effect. Student responses on the SET are subject to bias because students are influenced by instructor likability (Prasad et al., 2017). Studies conducted by McPherson et al. (2009) and Lawrence (2018) on leniency bias supported the argument that students rate instructors on the ease of the course, not on the effectiveness of the instruction.

More than 80% of student respondents indicated a preference for instructors who are assertive, clear, caring, and concerned for them outside of the classroom. The perception of these characteristics may be racial or gender-biased. For example, Wallace et al. (2019) found women of color are subject to stereotypical status and are evaluated by students as hostile and uncaring. MacNell et al. (2015) explained students subconsciously allow gender-role expectations, such as warmth and caring from female instructors and objectivity and assertiveness from male instructors, to influence their perceptions. Instructors who fail to display the expected gender role characteristics may experience negative comments on the SET from their students (Peterson et al., 2019).

Implications for Practice

The findings from this study support the suggestion made by Reinsch et al. (2020) that the SET should be used in conjunction with peer reviews and other forms of evaluation. One of the major concerns that emerged in the current study is the impact of implicit bias against non-native English-speaking instructors. The lack of student exposure to diverse faculty, paired with the preference for native English speakers, may result in negative evaluations for non-native English-speaking instructors. Educational institutions should construct SET items that minimize bias and provide more clarity. For

example, a question about the communication effectiveness of an instructor may result in lower ratings if students perceive an accent as a barrier to understanding.

As the student population becomes more diverse, students who identify as Non-White may expect colleges and universities to hire faculty who represent their racial and ethnic backgrounds. As reported in Chapter Four, over 30% of Latino students and 80% of Black/African American students stated they were more comfortable with instructors who look like them. The student experience of unrepresented minority groups can be enhanced by hiring faculty who are representative of these groups. In addition, students who identify as White would also benefit from interaction with instructors from other racial and ethnic groups.

Recommendations for Future Research

This study was limited to one department within a midwestern community college. Future research should include a larger sample from multiple departments, colleges, and universities. A mixed-methods approach would provide not only quantitative data but qualitative results to gain a broader perspective from faculty and student respondents. One limitation of this study was the small number of marginalized faculty respondents and the lack of faculty respondents identifying as African-American. Broadening the sample could yield perspectives from a more diverse group of faculty and students.

Interviews with participants from majority and minority groups would provide a greater understanding of how implicit bias influences perceptions of the SET. This current study was conducted in 100-level Communication and World Languages courses, and many of the student participants were first-semester freshmen. A study including students from upper-division courses is recommended, since those students have

experienced more college instruction and have completed the SET at the end of each semester.

Finally, a qualitative study focused on the impact of accents, dialects, and English as a second language would add to the research about implicit bias and the SET. As more studies are conducted about the impact of implicit bias on the SET, college communities will better utilize the information gained from the SET. The resulting effect is to improve teaching effectiveness without causing harm to marginalized faculty.

Summary

Chapter One was focused on the background of the study, theoretical framework, statement of the problem, and purpose. The purpose of this study was to measure the influence of implicit bias on the outcome of the SET at one Missouri community college. The variables examined in this study included gender, age, ethnicity, race, and English proficiency. The study also served as a guide to investigate how instructors perceive student interaction with the SET. This study was framed around the work of Greenwald and Banaji's (1995) theory of implicit bias, also known as implicit social cognition. Implicit bias is an unconscious mental process in which one perceives individuals or groups of people who are different in a negative way (Staats et al., 2015).

In Chapter Two, a review of literature was presented to provide an understanding of implicit bias and the history of the SET. The review included the strengths of the SET and concerns about SET validity, as noted by earlier researchers. The information in Chapter Two was focused on the impact of implicit bias in healthcare, the criminal justice system, employment, and education. Finally, the topics included in the chapter included information on marginalized groups such as women, underrepresented minorities, older instructors, and non-native English speakers.

In Chapter Three, the methodology utilized for collection and analysis was presented. This quantitative study was guided by three research questions. The sample was described, and the instruments were explained. The surveys were distributed and collected during the fall 2019 semester. Then, data were analyzed for each survey statement (Creswell & Creswell, 2017).

In Chapter Four, the results of the data were presented. The findings did not point to implicit bias based upon all of the characteristics examined. One finding was derived from the analysis of research question three; student responses indicated they evaluate instructors based on personality, friendliness, leniency, looks, or dress than teaching effectiveness. The literature reviewed in Chapter Two was consistent with these findings. A second finding was derived from the analysis of research question two. Student and faculty responses indicated a preference for instructors who speak English as their first language. This finding confirmed a degree of bias against instructors who speak English as their second language.

In Chapter Five, the findings were reported along with conclusions, implications for practice, and recommendations for future study. Even though the student and faculty responses did not indicate implicit bias based upon every characteristic examined, further research should be conducted using a larger and more diverse group of faculty and students. With nearly 90% of U.S. colleges and universities using the SET as a standard measure of instructor performance, it is imperative bias is identified and reduced (Wallace et al., 2019). For colleges and universities to recruit and retain marginalized faculty, they must respond to concerns surrounding the SET.

References

American Psychological Association *(2010). The publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Anderson, J., Brown, G., & Spaeth, S. (2006). Online student evaluations and response rates reconsidered. *Innovate: Journal of Online Education, 2*(6). Retrieved from https://nsuworks.nova.edu/innovate/vol2/iss6/5

Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavior Sciences, 2*, 184-201. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0739986304273707

Annamma, S., & Morrison, D. (2018). Identifying dysfunctional education ecologies: A discrit analysis of bias in the classroom. *Equity & Excellence in Education, 51*(2), 114-131. doi:10.1080/10665684.2018.1496047

Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles, 49*, 507-516. Retrieved from https://doi.org/10.1023/A:1025832707002

Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot hidden biases of good people.* New York, NY: Delacorte Press.

Basow, S. A., Codos, S., & Martin, J. L. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal*, *47*(2), 352-363. Retrieved from https://ldr.lafayette.edu/bitstream/handle/10385/1404/Basow-CollegeStudentJournal-vol47-2013.pdf?sequence=1

Bavishi, A., Madera, J. M., & Hebl, M. R. (2010). The effect of professor ethnicity and gender on student evaluations: Judged before met. *Journal of Diversity in Higher Education, 3,* 245-256. Retrieved from http://dx.doi.org/10.1037/a0020763

Berk, R. A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching and Learning*, *8*(2), 98-107. Retrieved from http://www.ronberk.com/articles/2012_top20.pdf

Bias. (2019). In *Oxford online dictionary*. Retrieved from https://en.oxforddictionaries.com/definition/bias

Bonitz, V. S. (2011). *Student evaluation of teaching: Individual differences and bias effects* (Doctoral dissertation, Iowa State University). Retrieved from https://lib.dr.iastate.edu/etd/12211

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Science Open Research.* doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review, 41*, 71-88. Retrieved from http://dx.doi.org/ 10.2139/ssrn.2004361

Calkins, S. C., & Micari, M. (2010). Less-than-perfect judges: Evaluating student evaluations. *Thought and Action, Fall 2010*, 7-22.

Campbell, J. P. (2005). *Evaluating teacher performance in higher education: The value of student ratings* (Doctoral dissertation, University of Central Florida). Retrieved from https://stars.library.ucf.edu/etd/438

Carlozzi, M. (2017). Rate my attitude: Research agendas and RateMyProfessor scores. *Assessment & Evaluation in Higher Education*, *43*(3), 359-368. doi:10.1080/02602938.2017.1348465

Causadias, J. M., & Umaña-Taylor, A. J. (2018). Reframing marginalization and youth development: Introduction to the special issue. *American Psychologist*, *73*(6), 707-712. Retrieved from https://doi.org/10.1037/amp0000336

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, *44*(5), 496-518. Retrieved from https://eric.ed.gov/?id=EJ676456

Clayson, D. E., & Haley, D. A. (2011). Are students telling us the truth? A critical look at the student evaluation of teaching. *Marketing Education Review, 15*(1), 1- 10. doi:10.1080/10528008.2005.11488884

Coladarci T., & Kornfield, I. (2007). Ratemyprofessors.com versus formal in-class student evaluation of teaching. *Practical Assessment, Research & Evaluation, 12*(6), 1-15.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement, 60*(6), 821-836. doi:10.1177//00131640021970934

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: Sage Publications.

Degheri, T. J. (2017). *An empirical look at the impact of course and faculty characteristics on student evaluations* (Doctoral dissertation, University of San Diego). Retrieved from http://digital.sandiego.edu/dissertations/76

Dehon, E., Weiss, N., Jones, J., Faulconer, W., Hinton, E., & Sterling, S. (2017). A systematic review of the impact of physician implicit racial bias on clinical decision making. *Academic Emergency Medicine*, *24*(8), 895-904. Retrieved from https://doi.org/10.1111/acem.13214

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. Retrieved from https://files.eric.ed.gov/fulltext/EJ1162267.pdf

Doubleday, A. F., & Lee, L. M. (2016). Dissecting the voice: Health professionals students' perceptions of instructor age and gender in an online environment and the impact on evaluations for faculty. *Anatomical Sciences Education*, *9*(6), 537-544. Retrieved from https://doi.org/10.1002/ase.1609

Echols, D. G., Neely, P. W., & Dusick, D. (2018). Understanding faculty training in competency-based curriculum development. *The Journal of Competency-Based Education, 3*(2). Retrieved from https://doi.org/10.1002/cbe2.1162

Fan Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R.,… Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLOS ONE, 14*(2). Retrieved from https://doi.org/10.1371/journal.pone.0209749

Feistauer, D., & Richter, T. (2018). Validity of students' evaluations of teaching: Biasing effects of likability and prior subject interest. *Studies in Educational Evaluation*, *59,* 168-178. Retrieved from https://doi.org/10.1016/j.stueduc.2018.07.009

Fiarman, S. E. (2016). Unconscious bias: When good intentions aren't good enough. *Educational Leadership, 74*(3), 10-15. Retrieved from https://www.educationalleadership-digital.com

Fink, A. (2016). *How to conduct surveys: A step by step guide* (6th ed.). Thousand Oaks, CA: Sage Publications, Inc.

FitzGerald, C., & Hurst, S. (2017). Implicit bias in healthcare professionals: a systematic review. *BMC Medical Ethics, 18* (19). Retrieved from doi:10.1186/s12910-017-8

FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019, December). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychology, 7*(1). Retrieved from doi:10.1186/s40359-019-0299-7

Flaherty, C. (2017, January 13). Study of online ratings of professors suggest scores vary with instructor's gender and perceived rigor. *Inside Higher Ed.* Retrieved from https://www.insidehighered.com/news/2017/01/13/study-online-ratings-professors-suggest-scores-vary-instructors-gender-and-perceived

Flaherty, C. (2018, July 2). Bye, bye, chili pepper: Rate my professors ditches its chili pepper "hotness" quotient. *Inside Higher Ed.* Retrieved from https://www.insidehighered.com/news/2018/07/02/rate-my-professors-ditches-its-chili-pepper-hotness-quotient

Floyd, K. (2018). *Communication matters*. New York, NY: McGraw-Hill Education.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed). New York, NY: McGraw-Hill.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27. Retrieved from https://faculty.washington.edu/agg/pdf/Greenwald_Banaji_PsychRev_1995.OCR.pdf

Gump, S. E. (2007). Student evaluations of teaching effectiveness and the leniency hypothesis: A literature review. *Educational Research Quarterly, 30*(3), 56-69.

Hamermesh, D. S., & Parker, A. M. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, *24*, 369-376. Retrieved from https://eric.ed.gov/?id=EJ697547

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, *4*(1). Retrieved from http://dx.doi.org/10.1080/2331186X.2017.1304016

Huston, T. A. (2006). Race and gender bias in higher education: Could faculty course evaluations impede further progress toward parity? *Seattle Journal for Social Justice*, *4*(2). Retrieved from https://digitalcommons.law.seattleu.edu/sjsj/vol4/iss2/34

Jackson, S. M. (2016). *The influence of implicit and explicit gender bias on grading, and the effectiveness of rubrics for reducing bias* (Doctoral dissertation, Wright State University). Retrieved from https://corescholar.libraries.wright.edu/etd_all/1529/

Jones, A. (2017). Implicit bias as social-framework evidence in employment discrimination. *University of Pennsylvania Law Review, 165*, 1221-1243. Retrieved from https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9581&context=penn_law_review

Khush, H. (2020, February 3). Guest opinion: Impacts of unconscious bias matter. *Business Record.* Retrieved from https://businessrecord.com/Content/Opinion/Article/Guest-opinion-Impacts-of-unconscious-bias-matter/168/963/89385

Kornell, N., & Hausman, H. (2016). Do the best teachers get the best ratings? *Frontiers in Psychology, 7*, 570. Retrieved from doi:10.3389/fpsyg.2016.00570

Kovera, M. B. (2019). Racial disparities in the criminal justice system: Prevalence, causes and a search for solutions. *Journal of Social Issues, 75*(4), 1139-1164. doi:10.1111/josi.12355

Kulik, J. A., & Kulik C. C. (1974). Student ratings of instruction. *Teaching of Psychology, 1*(2), 51-57. Retrieved from https://eric.ed.gov/?=Ej110077

Lawrence, J. W. (2018). Student evaluations of teaching are not valid. Retrieved from https://www.aaup.org/article/student-evaluations-teaching-are-not-valid

Legault, M. (2019, October 18). Implicit bias and disparate impact claims: A primer for employers (US). *The National Law Review, X*(344). Retrieved from https://www.natlawreview.com/article/implicit-bias-and-disparate-impact-claims-primer-employers-us

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation, 54*, 94-106. Retrieved from http://dx.doi.org/10.1016/j.stueduc.2016.12.004

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*, 291-303. doi:10.1007/s10755-014-9313-4

McClain, L., Gulbis, A., & Hays, D. (2018). Honesty on student evaluations of teaching: Effectiveness, purpose, and timing matter! *Assessment & Evaluation in Higher Education, 43*(3), 369-385. doi:10.1080/02602938.2017.1350828

McKeachie, W. K. (1990). Research on college teaching: The historical background. *Journal of Educational Psychology*, *82*(2), 189-200. Retrieved from https://pdfs.semanticscholar.org/e4bf/fe0109332da7d182acf6006458de07aae776.pdf

McPherson, M., & Jewell, R. T. (2007). Leveling the playing field: Should student evaluation scores be adjusted? *Social Science Quarterly*, *88*(3), 868-881.

McPherson, M., Jewell, T., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal*, *35*, 37-51.

Merino, Y., Adams, L., & Hall, W. J. (2018). Implicit bias and mental health professionals: Priorities and directions for research. *Psychiatric Services*, *69*(6), 723-725. Retrieved from doi.org/10.1176/appi.ps.201700294

Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *Political Science & Politics, 51*(3), 648-652. Retrieved from https://doi.org/10.1017/S104909651800001X

Murray, H. G. (2005, June). *Student evaluations of teaching: Has it made a difference*? Paper presented at the Annual Meeting of the Society for Teaching and Learning in Higher Education, Charlottetown, Prince Edward Island. Retrieved from https://www.stlhe.ca/wp-content/uploads/2011/07/Student-Evaluation-of-Teaching1.pdf

Nance, J. P. (2019). Implicit racial bias and students' fourth amendment rights. *Indiana Law Journal, 94*(1). Retrieved from https://www.repository.law.indiana.edu/ilj/vol94/iss1/2

Natfulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational deduction. *Journal of Medical Education*, *48*, 630-635. Retrieved from https://pubmed.ncbi.nlm.nih.gov/4708420/

National Center for Education Statistics. (2018). *National study of postsecondary faculty.* Retrieved from https://nces.ed.gov/surveys/nsopf/index.asp

National Center for Education Statistics. (2019). Integrated postsecondary education data system. Retrieved from https://nces.ed.gov/ipeds/

Oncidi, A. (2018, December 18). The coming battle over 'implicit bias' in employment discrimination cases. *Los Angeles & San Francisco Daily Journal.* Retrieved from https://s3.amazonaws.com/assets.production.proskauer/uploads/56f28d2eb05f2edba188ca4a7ce53a9b.pdf

Otani, K., Kim, B. J., & Cho, J. (2012). Student evaluation of teaching (SET) in higher education: How to use SET more effectively and efficiently in public affairs education. *Journal of Public Affairs Educations, 18*(3), 531-544. doi:10.1080/15236803.2012.12001698

Owen, A. (2019, June 24). The next lawsuits to hit higher education. *Inside Higher Education.* Retrieved from https://www.insidehighered.com/views/2019/06/24/relying-often-biased-student-evaluations-assess-faculty-could-lead-lawsuits-opinion

Payne, K., Niemi, L., & Doris, J. (2018). How to think about implicit bias. *Scientific American.* Retrieved from https://www.scientificamerican.com/article/how-to-think-about-implicit-bias/

Perry, A. R., Wallace, S. L., Moore, S. E., & Perry-Burney, G. D. (2014). Understanding student evaluations: A black faculty perspective. *Reflections: Narratives of*

Professional Helping, 20(1), 29-35. Retrieved from https://ir.library.louisville.edu/faculty/15

Peterson D. A. M., Biederman, L. A., Anderson, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS One*, *14* (5). Retrieved from doi:10.1371/journal.pone.0216241

Prasad, J. L., Ko, C. C., & Sanchez, A. (2017). Should student evaluation of teaching play a significant role in the formal assessment of dental faculty? Two viewpoints. Viewpoint 2: Student evaluation of teaching should not be part of formal faculty assessment. *Journal of Dental Education*, *81*(11). doi:10.21815/JDE.017.093

Price, J. H., & Payton, E. (2017). Implicit racial bias and police use of lethal force: Justifiable homicide or potential discrimination? *Journal of African American Studies*, *21*, 674-683. doi:10.1007/s12111-017-9383-3

Ray, V. (2018, February 9). Teaching evaluations are often used to confirm the worst stereotypes about women faculty [Editorial]. *Inside Higher Education*. Retrieved from https://www.insidehighered.com/advice/2018/02/09/teaching-evaluations-are-often-used-confirm-worst-stereotypes-about-women-faculty

Reid, L. (2010). The role of perceived race and gender in the evaluation of college teaching on ratemyprofessors.com. *Journal of Diversity in Higher Education*, *3*, 137-152. doi:10.1037/a0019865

Reinsch, R. W., Goltz, S. M., & Hietapelto, A. B. (2020). Student evaluations and the problem of implicit bias. *Journal of College & University Law*, *45*(1), 114-139. Retrieved from https://digitalcommons.mtu.edu/michigantech-p/1712/

Reints, R. (2018, July 11). After criticism, RateMyProfessor drops chili pepper rating that signaled a teacher's hotness. *Fortune.* Retrieved from https://fortune.com/2018/07/11/rate-my-professor-drops-chili-pepper-rating/

Reynolds, D. V. (1977). Faculty forum. *Teaching of Psychology*, *4*(2), 82-83.

Rodin, M., & Rodin B. (1972). Student evaluation of teaching. *Science, 177*(4055), 1164-1166. Retrieved from http://www.jstor.org/stable/1734252

Rosen, A. S. (2017). Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching: A large-scale study of ratemyprofessors.com data. *Assessment & Evaluation in Higher Education*, *43*(1), 31-44. doi:10.1080/02602938.2016.127615

Rowan, S., Newness, E. J., & Tetradis, S. (2017). Should student evaluation of teaching play a significant role in the formal assessment of dental faculty? Two viewpoints. Viewpoint 1: Formal faculty assessment should include student evaluation of teaching. *Journal of Dental Education*, *81*(11). doi:10.21815/JDE.017.093

Rynders, D. (2019). Battling implicit bias in the idea to advocate for African American students with disabilities. *Touro Law Review*, *35*(1). Retrieved from https://digitalcommons.tourolaw.edu/lawreview/vol35/iss1/18/

Sabin, J. A., & Greenwald, A. G. (2012). The influence of implicit bias on treatment recommendations for 4 common pediatric conditions: Pain, urinary tract infection, attention deficit hyperactivity disorder, and asthma. *American Journal of Public Health, 102*(5). doi:10.2015/AJPH.2011.300621

Sauer, T. (2012). *Predictors of student course evaluations* (Doctoral dissertation, University of Louisville). Retrieved from https://doi.org/10.18297/etd/1266

Sauermann, J., Mengel, F., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, *17*, 535-566. doi:10.1093/jeea/jvx057

Schmidt, P. (2015, March 26). Tenure lawsuit challenges privacy of student evaluations [Editorial]. *The Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/Tenure-Lawsuit-Challenges/228789

Schmidt, P. (2017, January 13). When students' prejudices taint reviews of instructors [Editorial]. *The Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/When-Students-Prejudices/238892

Smith, B. P. (2007). Student ratings of teaching effectiveness: An analysis of end of course faculty evaluations. *College Student Journal*, *41*, 788-800. Retrieved from https://eric.ed.gov/?id=EJ816803

Smith, B. P., & Hawkins, B. (2011). Examining student evaluations of black college faculty: Does race matter? *The Journal of Negro Education*, *80*(2), 149-162.

Spencer, K. B., Charbonneau, A. K., & Glaser, J. (2016). Implicit bias and policing. *Social and Personality Psychology Compass*, *10*(1), 50-63. Retrieved from doi.org/10.1111/spc3.12210

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*, 598-642. Retrieved from https://doi.org/10.3102/0034654313496870

Spooren P., & Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and their relationships with SET scores. *Studies in*

Educational Evaluation, *54*, 43-49. Retrieved from http://dx.doi.org/10.1016/j.stueduc.2016.12.003

Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, *53*(11/12). doi:10.1007/s11199-005-8292-4

Sproule, R. (2000). Student evaluation of teaching: A methodical critique of conventional practices. *Educational Policy Analysis Archives, 8*(50) 1-23. doi:10.14507/epaa.v8n50.2000

Staats, C. (2016). Understanding implicit bias what educators should know. *American Educator, Winter*(2015-2016), 29-33. Retrieved from https://www.aft.org/ae/winter 2015-2016/staats

Staats, C., Capatosto, K., Tenney, L., & Mamo, S. (2017). *State of the science: Implicit bias review 2017*. Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity. Retrieved from http://kirwaninstitute.osu.edu/implicit-bias-training/resources/2017-implicit-bias-review.pdf

Staats, C., Capatosto, K., Wright, R. A., & Contractor, D. (2013). *State of the science: Implicit bias review 2013*. Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity. Retrieved from http://kirwaninstitute.osu.edu/wp-content/uploads/2015/05/2015-kirwan-implicit-bias.pdf

Staats, C., Capatosto, K., Wright, R. A., & Contractor, D. (2015). *State of the science: Implicit bias review 2015*. Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity. Retrieved from http://kirwaninstitute.osu.edu/wp-content/uploads/2015/05/2015-kirwan-implicit-bias.pdf

Stalnaker, J. M., & Remmers, H. H. (1928). Can students discriminate traits associated with success in teaching? *Journal of Applied Psychology*, *12*(6), 602-610. Retrieved from http://arldocdel.iii.com/1457557.pdf

Stark, P. B., & Freishatat, R. (2014). An evaluation of course evaluations. *Science Open Research.* doi:10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Stevens, C. M., Schneider, E., & Bederman-Miller, P. (2018). Identifying faculty perceptions of awareness and preparedness relating to ADA compliance at a small, private college in NE PA. *American Journal of Business Education*, *11*(2), 27-40. Retrieved from https://files.eric.ed.gov/fulltext/EJ1175335.pdf

Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. *Perspectives on Psychological Science*, *11*(6), 800-816. Retrieved from https://doi.org/10.1177/1745691616650284

Thielschi, M., Brinkmoller, B., & Forthmann, B. (2018). Reasons for responding in student evaluation of teaching. *Studies in Educational Evaluation*, *56,* 189-196. Retrieved from https://doi.org/10.1016/j.stueduc.2017.11.008

Thijs, J., & Verkuyten, M. (2016). Ethnic attitudes and social projection in the classroom. *Child Development*, *87*(5), 1452-1465. Retrieved from https://doi.org/10.1111/cdev.12597

Türko, E. S. (2016). Can entrepreneurship education reduce stereotypes against women entrepreneurship? *International Education Studies*, *9*(11), 53-65. Retrieved from https://doi.org/10.5539/ies.v9n11p53

Tyner, A. (2019). Unconscious bias, implicit bias, and microaggressions: What can we do about them? *GPSolo.* Retrieved from

https://www.americanbar.org/groups/gpsolo/publications/gp_solo/2019/july-august/unconscious-bias-implicit-bias-microaggressions-what-can-we-do-about-them/

U.S. Census Bureau. (2017). *Race & ethnicity.* Retrieved from https://www.census.gov/mso/www/training/pdf/race-ethnicity-onepager.pdf

U.S. Census Bureau. (2018). Gender. In *Census.gov glossary.* Retrieved from https://www.census.gov/glossary/#term_Gender

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22-42. Retrieved from https://doi.org/10.1016/j.stueduc.2016.08.007

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education, 23*(2), 191-212. doi:10.1080/0260293980230207

Wallace, S. L., Lewis, A. K., & Allen, M. D. (2019). The state of the literature on student evaluations of teaching and an exploratory analysis of written comments: Who benefits most? *College Teaching*, *67*(1), 1-14. doi:10.1080/87567555.2018.1483317

Whitfield, M. (2019). *Influence of implicit-bias training on the cultural competency of police officers* (Doctoral dissertation, Walden University). Retrieved from https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=8374&context=dissertations

Young, K., Joines, J., Standish, T., & Gallagher, V. (2018). Student evaluations of teaching: The impact of faculty procedures on response rates. *Assessment &*

Evaluation in Higher Education, *44*(1), 37-49.

doi:10.1080/02602938.2018.1467878

Appendix A

Student Survey

Thank you for agreeing to complete this survey. The findings will be used for academic research to explore the influence of implicit bias on student evaluations of teaching. The survey should take approximately 5-7 minutes to complete.

Please respond to the following:

My gender: ____Male ____Female

My age: _____18-25 _____ 26-40 _____ 41-55 _____56+

My race: _____ White

_____ Black/African American

_____ Hispanic/Latino

_____ Native American/American Indian

_____ Asian/Pacific Islander

_____ Other

1. By evaluating my professors, I am actually helping them improve their teaching.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

2. The course evaluation form is adequate to evaluate my professors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

3. Students should take student evaluations of teaching seriously.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

4. I read and understand each statement before I rate it.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

5. I prefer taking courses from male instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

6. I prefer taking courses from female instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

7. I prefer taking courses from young and enthusiastic instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

8. I prefer taking courses from older and experienced instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

9. I am comfortable taking courses from instructors who speak English as their second language.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

10. I am more comfortable taking courses from instructors who speak English as their first language.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

11. When evaluating professors, I pay more attention to their personalities (e.g., friendliness, leniency, looks, or dress) than teaching effectiveness.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

12. I prefer instructors who are assertive and provide clear expectations for the course.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

13. I prefer instructors who are caring and concerned about me outside of the classroom.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

14. I have taken a class from an instructor of a different race or ethnicity.
    - Agree
    - Disagree

15. I am excited to take a class from an instructor from a different race or ethnicity than my own.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

16. I am more comfortable with instructors who appear to be from my own racial or ethnic background.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

Appendix B

Faculty Survey

Thank you for agreeing to complete this survey. The findings will be used for academic research to explore the influence of implicit bias on student evaluations of teaching. The survey should take approximately 5-7 minutes to complete.

Please respond to the following:

My gender: ____Male ____Female

My age: _____25-35 _____ 36-50 _____ 51-64 _____65+

My race: _____ White

_____ Black/African American

_____ Hispanic/Latino

_____ Native American/American Indian

_____ Asian/Pacific Islander

_____ Other

1. The course evaluation helps improve my teaching effectiveness.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

2. The course evaluation form is adequate to evaluate my teaching effectiveness.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

3. Students should take student evaluations of teaching seriously.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

4. Students read and understand each statement before they rate it.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

5. Students prefer taking courses from male instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

6. Students prefer taking courses from female instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

7. Students prefer taking courses from young and enthusiastic instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

8. Students prefer taking courses from older and experienced instructors.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

9. Students are comfortable taking courses from instructors who speak English as their second language.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

10. Students are more comfortable taking courses from instructors who speak English as their first language.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

11. Students pay more attention to instructor personality (e.g., friendliness, leniency, looks, or dress) than teaching effectiveness.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

12. Students prefer instructors who are assertive and provide clear expectations for the course.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

13. Students prefer instructors who are caring and concerned about them outside of the classroom.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

14. Many students have taken a class from an instructor of a different race or ethnicity.
    - Agree
    - Disagree

15. Many students are excited to take a class from an instructor from a different race or ethnicity than their own.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

16. Many students are more comfortable with instructors who appear to be from their own racial or ethnic background.
    - Strongly agree
    - Agree
    - Neutral
    - Disagree
    - Strongly disagree

Appendix C

Permission Letter

BERRY, KIMBERLY F.

From: BERRY, KIMBERLY
Sent: Thursday, November 15, 2018 4:45 AM
To: Smitha Dev
Subject: Formal request to use survey

Greetings Dr. Dev:

Thank you for your response about your article, "Major Factors Affecting Students' Perception Towards Faculty Evaluation of Teaching" found in the *Journal of Social Studies Education Research*, 8(3) 149-167. This email is a official request to use statements from your survey as my instrument in efforts to complete the research requirement for an Ed.D. at Lindenwood University, St. Charles, Mo. USA. The survey will be distributed to approximately 2,000 college freshman and sophomores at a local community college in Missouri. Thank you for your willingness to assist me with this endeavor.

Regards,

Kimberly Berry

Kimberly F. Berry, M.A.
Doctoral Candidate
Lindenwood University
St. Charles, Missouri

From: Smitha Dev <Smitha.Dev@adu.ac.ae>
Sent: Sunday, November 18, 2018 1:52:23 AM
To: BERRY, KIMBERLY
Subject: RE: Formal request to use survey

Hi Mr. Berry,

This is to inform you that I don't have any objection to you in using our *Students' Perception Towards Faculty Evaluation of Teaching scale* in your PhD thesis work.

Regards

Dr. Smitha Dev
Assistant Professor of Psychology
General Psychology Course Coordinator, University College

Abu Dhabi University,
P.O. Box 59911
Abu Dhabi, UAE

T: +971 2 5015627, F: +971 2 5860176

www.adu.ac.ae

Follow us on:
/AbuDhabiUni

Appendix D

IRB Approval

Nov 11, 2019 5:36 PM CST

RE:
IRB-20-86: Initial - The Influence of Implicit Bias on Student Evaluations of Teaching at a Missouri Community College

Dear Kimberly Berry,

The study, The Influence of Implicit Bias on Student Evaluations of Teaching at a Missouri Community College, has been Approved as Exempt.

Category: Category 1. Research, conducted in established or commonly accepted educational settings, that specifically involves normal educational practices that are not likely to adversely impact students' opportunity to learn required educational content or the assessment of educators who provide instruction. This includes most research on regular and special education instructional strategies, and research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

The submission was approved on November 11, 2019.

Here are the findings: **Regulatory Determinations**

- This study has been determined to be minimal risk because the research is not obtaining data considered sensitive information or performing interventions posing harm greater than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.
  Sincerely,
  Lindenwood University (lindenwood) Institutional Review Board

Appendix E

Letter of Participation for Students

Date:

Students,

This survey is part of research conducted through Lindenwood University by Kimberly Berry under the guidance of Dr. Sherry DeVore. The purpose of this study entitled *The Influence of Implicit Bias on Student Evaluations of Teaching at a Missouri Community College* is to examine student feedback of instructor performance.

Your participation in this research will involve the completion of an online survey that will take approximately 5-7 minutes.

All students enrolled in a course in the Communication and World Languages department are invited to participate in this research (this is approximately 2,000 potential participants).

An informed consent form is available for you to read before agreeing to participate in this study.

Thank you in advance for your time and assistance!

Appendix F

Letter of Participation for Faculty

Date:

Hello [Name of Potential Respondent],

This email is an invitation for you to participate in a research study conducted through Lindenwood University by Kimberly Berry under the guidance of Dr. Sherry DeVore. The purpose of this study entitled *The Influence of Implicit Bias on Student Evaluations of Teaching at a Missouri Community College* is to examine student feedback of instructor performance.

Your participation in this research will involve the completion of an online survey that will take approximately 5-7 minutes. All Communication and World Languages faculty are invited to participate in this research (this is approximately 60 potential participants). An informed consent form is available for you to read before agreeing to participate in this study.

Thank you in advance for your time and assistance!

Appendix G

LINDENWOOD

Survey Research Information Sheet
Student

You are being asked to participate in a survey conducted by Kimberly Franklin Berry and Dr. Sherry DeVore at Lindenwood University. We are conducting this study to explore how instructor characteristics such as ethnicity, race, age, gender, accent, and personality influence the outcome of the SET at a Missouri community college. The study will also serve as a guide to investigate whether implicit bias influences how students respond to the SET.

It will take about 5-7 minutes to complete this survey.

Your participation is voluntary. You may choose not to participate or withdraw at any time by simply not completing the survey or closing the browser window.

There are no risks from participating in this project. We will not collect any information that may identify you. There are no direct benefits for you participating in this study.

WHO CAN I CONTACT WITH QUESTIONS?

If you have concerns or complaints about this project, please use the following contact information:

Kimberly Franklin Berry KFB447@lionmail.lindenwood.edu
Dr. Sherry DeVore sdevore@lindenwood.edu

If you have questions about your rights as a participant or concerns about the project and wish to talk to someone outside the research team, you can contact Michael Leary (Director - Institutional Review Board) at 636-949-4730 or mleary@lindenwood.edu.

By clicking the link below, I confirm that I have read this form and decided that I will participate in the project described above. I understand the purpose of the study, what I will be required to do, and the risks involved. I understand that I can discontinue participation at any time by closing the survey browser. My consent also indicates that I am at least 18 years of age.

You can withdraw from this study at any time by simply closing the browser window. Please feel free to print a copy of this information sheet.

Appendix H

LINDENWOOD

Survey Research Information Sheet
Faculty

You are being asked to participate in a survey conducted by Kimberly Franklin Berry and Dr. Sherry DeVore at Lindenwood University. We are conducting this study to explore how instructor characteristics such as ethnicity, race, age, gender, accent, and personality influence the outcome of the SET at a Missouri community college. The study will also serve as a guide to investigate whether implicit bias influences how students respond to the SET.

It will take about 5-7 minutes to complete this survey.

Your participation is voluntary. You may choose not to participate or withdraw at any time by simply not completing the survey or closing the browser window.

There are no risks from participating in this project. We will not collect any information that may identify you. There are no direct benefits for you participating in this study.

WHO CAN I CONTACT WITH QUESTIONS?

If you have concerns or complaints about this project, please use the following contact information:

Kimberly Franklin Berry KFB447@lionmail.lindenwood.edu
Dr. Sherry DeVore sdevore@lindenwood.edu

If you have questions about your rights as a participant or concerns about the project and wish to talk to someone outside the research team, you can contact Michael Leary (Director - Institutional Review Board) at 636-949-4730 or mleary@lindenwood.edu.

By clicking the link below, I confirm that I have read this form and decided that I will participate in the project described above. I understand the purpose of the study, what I will be required to do, and the risks involved. I understand that I can discontinue participation at any time by closing the survey browser. My consent also indicates that I am at least 18 years of age.

You can withdraw from this study at any time by simply closing the browser window. Please feel free to print a copy of this information sheet.

Vita

Kimberly Franklin Berry was born in Springfield, Missouri, and attended public school at Phelps Elementary, Pipkin Jr. High, and Central High School. Kimberly continued her post-secondary education at Southwest Missouri State University (SMSU) and graduated in 1987 with a Bachelor of Science degree in electronic media and a minor in journalism. She then pursued a Master of Arts in Communication from SMSU, earning her degree in 1992.

During her Master's work, Kimberly was employed at Southwest Missouri State University as a minority recruiter for the Admissions Office. She also served as an academic advisor and taught college reading and study skills at the university. After the completion of her Master's degree, Kimberly began teaching as an adjunct at a community college. In the fall of 1997, Kimberly accepted a full-time faculty position in the Communication Department at the college. In 2007, Kimberly became the department chair for the Communication Department. In 2010, the department expanded to include foreign languages. Kimberly currently serves as the chair of the Department of Communication and World Languages.