Electronic Theses and Dissertations

8-2018

Examining the Meaning of Course Evaluation

Tyesha Stewart
Stephen F. Austin State University, stewarttd1@jacks.sfasu.edu

Follow this and additional works at: https://scholarworks.sfasu.edu/etds

Part of the Academic Advising Commons, Curriculum and Instruction Commons, Education Policy Commons, Gender and Sexuality Commons, Higher Education Administration Commons, Higher Education and Teaching Commons, Online and Distance Education Commons, Race and Ethnicity Commons, Scholarship of Teaching and Learning Commons, and the School Psychology Commons

Tell us how this article helped you.

Examining the Meaning of Course Evaluation

Creative Commons License

EXAMINING THE MEANING OF COURSE EVALUATION

By

Tyesha De'Shuan Stewart, B. A.

Presented to the Faculty of the Graduate school of

Stephen F. Austin State University

In Partial Fulfillment

Of the Requirements

For the Degree of

Master of Arts

STEPHEN F. AUSTIN STATE UNIVERSITY

August, 2018

EXAMINING THE MEANING OF COURSE EVALUATION

By

Tyesha De'Shuan Stewart, B. A.

APPROVED:

_____
Robbie Steward, Ph. D., Thesis Director

_____
Luis Aguerrevere, Ph. D., Committee Member

_____
Nina Ellis-Hervey, Ph. D., Committee Member

_____
Sarah Savoy, Ph. D., Committee Member

_____
Pauline Sampson, Ph. D.
Dean of Research and Graduate Studies

Abstract

Teaching evaluations significantly influence faculty members' annual evaluations and progress toward tenure and promotion within academic programs in higher education. The purpose of this study is to add to the literature by providing a more in-depth examination of the meaning of course evaluations. Analyses were performed to address the following research question: "To what degree do faculty demographics, academic background, academic status, professional engagement, course grading patterns, and course characteristics predict the variance in overall course evaluation, standard deviation, and response rates?" Collected data consisted of course evaluations of 286 Human Services-related courses and information from 55 faculty members' vitae, during one academic year in an institution. Analysis of this data indicated that faculty academic background significantly predicted the variance in overall course evaluations and standard deviations. Results from this study may assist with policy development regarding promotion and tenure and more effective procedures for the evaluation of course instruction.

Table of Contents

List of Tables

Chapter I

Introduction

University course evaluations have been a common and pervasive practice within higher education settings (Chen & Hoshower, 2003). The measurements of teaching effectiveness allow students the opportunity to provide vital feedback for the purpose of improving the quality of learning in courses and teaching performance. After publishing the first teacher rating scale in 1915, this measurement was then introduced in the 1920s across several universities in the United States (Wachtel, 1998). Course evaluation surveys were distributed to students at the end of every semester to inform the instructors and those assigned the responsibility of evaluating instructors information about students' perception of course content, the instructor's manner in the classroom, and teaching methods (Smith & Welicker-Pollak, 2008). The electronic or paper-format distribution and collection of the surveys occur in a manner to protect students' anonymity and is composed of Likert-style items that allow students an opportunity to rate the course and instructor at the end of an academic period (Dommeyer, Baum, Hanna, & Chapman, 2004). Many also include a section wherein students might provide a narrative response about the quality of the course that extends beyond the items. Student participation provides insight to faculty members about areas of strength and areas that may need to be changed or improved within the course to enrich the learning environment (Ballantyne, Borthwick & Packer, 2000; Boysen, 2016; Dunrong & Fan 2009; Marsh & Roche, 1997).

Course evaluations are viewed as one of the most popular and meaningful measurements in providing necessary information concerning teaching effectiveness, as this instrument becomes an important tool adopted across various universities worldwide (Ahmadi, Helms, & Raiszadeh, 2001; Chen & Hoshower, 2003; Hammonds, Mariano, Ammons, & Chambers, 2017; Hobson & Talbot, 2001; Malouff, Reid, Wilkes, & Emmerton, 2015). The evaluations not only provide feedback to the faculty members to improve instructional performance, but also provide information to the administration during the decision of faculty members' promotion and tenure process. As this tool is constantly being utilized for the purpose of self-improvement and in determination of 'fit' for the position as an academician, debate about the validity and reliability of commonly used rating scales among both on the evaluation of teaching effectiveness (Clayson, 2009; Nasser & Fresko, 2002).

With faculty promotion and tenure weighing so heavily on course evaluation, the results of course evaluations are being questioned and challenged (Spooren, Brockx, & Mortelmans, 2013). This is especially a current and most relevant issue given the persisting under-representation of racial and ethnic minorities who experience upward mobility as academicians within predominantly White universities and colleges. Anecdotal reports (Armstrong, 1998; Eiszler, 2002; Sproule, 2000) and empirical findings purport a potential influence of students' biases and pre-conceived expectations regarding varying points of diversity. Combined with implications of decision-making with serious outcomes being based on the ratings of untrained evaluators, examination of

the variables that influence the outcome would seem imperative. Nevertheless, that the body of literature associated with the empirical examination of course evaluation ratings is limited (Beran & Rokosh, 2009), and in cases wherein investigations do occur, the scope is restricted to only a few variables (i.e. course level, class size, course grade). Students' ratings have yet to be comprehensively examined, and our practices remain based on the assumption that course evaluation ratings are accurate reflections of instructors' effectiveness. For this reason, there rise several concerns regarding course evaluation being used as a success-based tool, including untrained evaluators and the criteria used to define effective teaching (Giovannelli, 2003).

Success Based Tool

The course evaluation system is merely viewed as a success-based tool that contributes significantly to faculty promotion and tenure because it measures faculty members' quality in instruction (Chan, Luk, & Zeng, 2014; Emery, Kramer, & Tian, 2003; McPherson & Jewell, 2007; Simpson, & Siguaw, 2000). The various measures used to evaluate course instruction vary from institution to institution. The items address topics such as overall rating of course content, consistency in coursework, course difficulty, fairness in grading scale, available resource, knowledge of content, presentation quality, faculty member enthusiasm & flexibility, faculty member attitude and interaction with the student, and encouragement towards the student (Abrami, D'Apollonia, & Rosenfield, 2007; Ginns, Prosser, & Barrie, 2007). Students' learning outcomes and degree of interest in the content are also noted. Some measures request

students to note their anticipated grade to at least check if these variables might contribute to the overall rating (Filak & Sheldon, 2003). However, even in these cases, faculty seldom receive feedback about significant relationships found and are left to their own to identify and explain student variables that may influence their ratings (i.e., student academic ability, student investment in the class, students' race/ethnicity). Also, the questionnaire goes beyond asking about the course content by including domain items about the faculty member's demeanor, personality, and the general student's satisfaction with course (Denson, Loveday, & Dalton, 2010). For example, questions such as 'the instructor's enthusiasm' or 'how satisfied were you with this course' are questions listed outside of the course content. According to Tagomori and Bishop (1995), these items within the rating scale are argued to be ambiguous, unclear, and skewed with no necessary purpose in regards to effective teaching. It has been concluded by some that seemingly tangential, vaguely defined constructs (i.e., satisfaction) and student-controlled questions (i.e., perceptions of learning, interest in content) included asked in the evaluation, results in subjective responses from students which affect the overall score of the quality of faculty teaching (Uttl, White, & Gonzalez, 2017).

Personal satisfaction of the course from student evaluators becomes a critical issue for faculty seeking promotion and tenure and toward the overall rating of the course evaluation due to a focus on students' interest level, amount of information learned, and if students received the help and accommodations as requested from their faculty member (Titus, 2008). These evaluations are more likely to result in a sense of entitlement from

students, lower academic standards, and grade inflation (Cain, Romanelli, & Smith, 2012). The researchers express concerns with students believing and expecting they should receive a high grade without investing effort in academic performance. Gray and Bergmann (2003) mention that some faculty members oppose the use of the course evaluation tool because of the inaccurate reflection of their teaching performance. From the instructors' perspective, a rating scale may not measure all the qualities that influence effective teaching. The tool may be used improperly, and effective teaching and course requirements may be reduced in favor of the student to gain better scores. Also, there is a possibility of decrease in job satisfaction due to the critical feedback provided by students of instructors teaching method. For example, the tool may be misused if the evaluation feedback places more concern on students' level of satisfaction with the teaching than the students' perceptions of the quality of teaching. Instead, there should be a limited amount of questions provided in regards to students' satisfaction, and a greater focus on the quality of teaching (Richardson, 2005). Faculty members may be more likely to give easier assignments, reduce students' workload, and participate in grade inflation to obtain a higher course evaluation. To ensure that course evaluations are an accurate reflection of quality of teaching, all faculty members' expectations of their course should align with their university or department standards (Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Stroebe, 2016). The lack of this method could be due to an unrealistic or unwarranted goal, that most faculty may have difficulty following. Also, a low rating on student evaluations may reduce faculty members' job satisfaction and effort to be

engaged in creative instructional methods out of fear of students' responses. Satisfying

students to obtain a higher course evaluation then becomes a more primary objective than

teaching and challenging students to the best of their ability (Algozzine et al., 2004).

Unlike the close-ended Likert-style items, the open-ended questions are found to

be attended to less because it is qualitative data and difficult to analyze than quantitative

data collected on course evaluations (Alhija & Fresko, 2009; Lewis, 2001; Sheehan &

DuPrey, 1999). Nonetheless, current research has found text and data mining to be an

effective method to analyze written responses as a result of technology advances (Chen &

Chen, 2009; Romero, Ventura, & García, 2008; Stupans, McGuren, & Babey, 2016;

Thomas, 2014). Text mining is the process of identifying the relationships between texts

to interpret patterns, frequency, and relationships with other words (Tucker, 2014). The

narrative response provided by students allow them a chance to give commentary on

specific teaching tactics or content they enjoyed, and mention issues the items on the

evaluation does not address. Also, the students have an opportunity to provide specific

recommendation for each individual instructor so they may improve their course and

teaching strategies (Hodges & Stanton; 2007; Wongsurawat, 2011). Yet, the concern

raised by faculty is how constructive students' commentary may be to their evaluation

and if the comments provided are unwarranted. In some cases, students' comments can

be found to be quite offensive and unprofessional and thus becomes invaluable to faculty

(Tucker, 2014).

The format of the course evaluation system seems to address the purpose of giving valuable information to improve teaching performance, but it overshadows how effective this tool can be when introducing questions to students that suggest providing a subjective perspective of their faculty member. Specifically, subjective questions on course evaluations refer to students' self-assessment of their learning experience and their personal interpretation of how they would rate their faculty member. For example, questions such as 'the amount I learned during the course' or 'my interest level of each class session' are open to interpretation instead of being based on actual learning outcomes. Students' learning, of course, is and should be associated with the experience in the classroom, however, it also is influenced by variables such as prior academic preparation, study habits, class attendance, and consistent completion of assignments based on criteria established in course syllabi. These variables are not considered in most commonly used measures. The relevance of the inclusion of the currently used questions without additional inquiry regarding student variables is currently solely based on the students' perspective, and some have concluded that this long-standing practice should be subject to greater scrutiny (Berk, 2005; Simpson & Siguaw, 2000; Stehle, Spinath, & Kadmon, 2012).

Untrained Evaluators

What is most important to note in this section is that college/university students have no prior experience in evaluating instruction. Not only are they untrained, but they are without experience in rating instructors. It is the exception that there is such

opportunity during K-12 and given that the ability to think critically has been identified

as a deficit among college graduates by employers in the work force (Flores, Matkin,

Burbach, Quinn, & Harding, 2012), it becomes even more imperative that researchers are

recommended to more extensively examine the variables that may significantly influence

students' ratings to better understand the potentially life-changing outcomes in terms of

promotion, tenure, and merit among the most highly educated within society. The

prevailing question is: How might one be an effective and reliable evaluator of another's

performance if critical thinking competence, experience, and training are limited?

　　　In relevance to entering the profession as a School Psychologists, the traditional

perception of their work has been restricted to grades PK-12, but it extends beyond into

higher education settings (Sandoval & Love, 1977). Their contributions within the school

setting to promote and support students' ability to learn are link to college preparation

and readiness. Their expertise in learning and behavior allow them to assist students with

developing critical thinking and adjust to learning environments. It is an unfair act to

expect students to score or rate something they do not fully understand. Early exposure of

evaluations may help these students comprehend similar tactics used in college, and

possibly gain insight of what is expected of them as they enter higher education. For

School Psychologists who choose faculty appointment in higher education settings, this

information is essential to understanding how to interpret and use valuable information

provided from evaluations. School Psychologists are trained in areas such as data

collection and analysis, progress monitoring, academic interventions, instructional

support, diversity in development and learning, and research and program evaluation which corresponds significantly to course evaluations and teaching performance (NASP, 2017). This tool may help faculty members in the program of school psychology acknowledge whether student learning outcomes and content domains are being met and addressed according to the National Association of School Psychologists (NASP) as they prepare graduate students entering the profession.

After the completion of each course, each student is given the opportunity to evaluate their faculty member and the overall quality of the class based on their personal experience. Each evaluation is anonymous and confidential, allowing students to be honest and objective as they critique with the focus primarily on the course content, instructor's knowledge, and content presentation. But is objectivity really ever identified as a responsibility? Faculty members expect their students to be truthful when providing evaluations though there is no requirement or control to fulfill this obligation. Concerns of students' bias in teaching evaluations exist because assumptions arise that students are not consistently honest in their feedback (Clayson & Haley, 2011). It is this very fact that supports questions about the 'untrained evaluator'. That is also the reason that some institutions request that students indicate the anticipated final grade so that reviewers might be able to determine if there is a relationship between the rating and the student's performance in the course (Crumbley, Henry, & Kratchman, 2001). However, this is not a consistent practice across institutions or even programs. Even in cases wherein this information is collected, seldom is it considered in reviews for promotion and tenure.

Students may also believe that faculty will be able to determine the rater and that this info will affect future interaction and grading. This is especially the case in graduate-level courses wherein students will presumably have multiple course enrollments with the same instructor. The concern is faculty retribution for students' negative commentary and ratings in prior courses. Some students may believe that their voices will not be considered so will not respond; others may not see the rating as an important process and neglect to submit responses for no reason at all, being distracted by other competing life events and responsibilities. There are also some students who may not be aware or informed of the implications of no response, which decreases the validity of overall course evaluation, especially in cases of low response rates.

While completing this survey numerous variables may influence students' ratings other than questions directed only towards effective instruction. The characteristics of the instructor (i.e. experience, reputation, research involvement, personality, gender, race, physical appearance), student (i.e. personality, subject interest, gender, race, age, emotional state, expectation, study skills, level of academic preparation, family/life responsibilities, work commitment), course (i.e. meeting time, level, class size, subject area, rigor of the instruction and assignments), and administration of student evaluations (i.e. timing, anonymity of student raters, instructor potential access to the individual ratings, purpose of evaluation) may have substantial impact in influencing student scores of teaching effectiveness (Onwuegbuzie et al., 2017).

If course evaluation is purely based on students' point of view, there is likely to be a greater chance of subjective opinions influencing scores and response rates which stress another issue regarding untrained evaluators. Students' experiences are valid in understanding the instructional setting and the instructor but should not be collected without considering who the raters are in relationship to the instructor for the same reasons we carefully monitor other testing scenarios. Faculty members have voiced concerns of students not being qualified to evaluate their course and teaching effectiveness (Emery et. al, 2003). One recommendation is required training for students during orientation prior to assigning the critical responsibility of rating teaching effectiveness due to the degree of influence it has on faculty academic status. However, in a review of the literature, no such practice was noted.

Though course evaluation is mentioned within most course syllabi for quality feedback to improve the class, there is no guideline to ensure students are rating faculty teaching effectiveness with a sincere and serious attitude, and without allowing their own personal bias to affect their response. Studies have shown multiple factors to influence student ratings such as attractiveness to teacher personality, academic discipline, course level, and student achievement within the class (Emery et. al, 2003; Benton & Cashin, 2012). It would be more accurate to identify empirically supported variables associated with teaching effectiveness at the college level and include only these items (Stark & Freishtat 2014). Also, it would be more accurate to designate a section of items that

address student self-evaluation questions but are not included in the rating of the faculty member.

The validity of ratings may also be influenced by student's response rates. For example, the meaningfulness of ratings may be negatively affected by students' motivation to complete course evaluations, especially when there is no necessary requirement for every student to partake in taking the survey to submit their feedback. Furthermore, students may lack the awareness of how important their feedback in course evaluation is to faculty members and choose not to participate in course evaluation unless influenced to complete such evaluation due to the use of incentives (Goodman, Anson, Belcheir, 2015; Jaquett, VanMaaren, & Williams, 2016). However, the inclusion of incentives for this purpose presents a concern for some faculty that is the same in cases of research participation associated with some form of incentive (Goodman et al., 2015; Jaquett et al., 2016). The limitations of using incentives may result in grade inflation and decrease the validity of faculty members' teaching effectiveness (Love & Kotchen, 2010). There is also the concern about establishing a practice of offering incentives for that which is a noted as a student responsibility in many program handbooks.

Criteria of Effective Teaching

When reviewing faculty members' dossiers for promotion and tenure, responsibilities and contributions to their department within the university are reviewed. These items may include assignments in advising students, research participation, service to the school and community, presentations, publications, professional license and

certifications, and teaching effectiveness. The course evaluation is used as a representation of the system to measure teaching effectiveness based on students' feedback of the course, which is also used as part of the promotion and tenure process (Baldwin & Blattner, 2003; Subbaye & Vithal, 2017; Zabaleta, 2007). In university and college settings designated as teaching institutions, the weight of these course evaluations in the review process is even greater.

The concern with this system is how the supervisor and committee members define the criteria for effective teaching and whether the course evaluation is a sufficient measurement of teaching performance. Subbaye and Vithal (2017) conducted a study to identify which criteria significantly influence overall evaluation of teaching and academic promotion. The following teaching criteria were listed for evaluating academic promotion: rationale for approach to teaching, teaching methods, postgraduate supervision, assessing students' work and performance, peer evaluations, student evaluations, ongoing study of tertiary education, development of new courses, sharing teaching experience with others, and special recognition of teaching. Though the criteria are not mandatory across every rank, they are each specifically important for review across the rank levels of lecturer, senior lecturer, associate faculty member and full faculty member. Postgraduate supervision and sharing teaching experiences were statistically significant criteria in overall evaluation of teaching and promotion. On the contrary, some studies reported research productivity and student evaluations as the most common criteria for effective teaching and promotion (Parker, 2008; Chen &Yeager,

2011). The reason for the difference among these studies is because there is not one specific criteria tool that is used extensively across institutions for review. There is a variety of criteria being used by institutions to evaluate teaching effectiveness and academic promotions, and among these selected criteria only specific items apply to faculty rank when submitting a teaching portfolio for evaluation.

Despite many arguments against student evaluations (Boring, Ottoboni, & Stark, 2016; King & Fraser, 2016, Milanowski, 2017), it has been proven to be an effective measurement of instructors' teaching effectiveness because of positive correlation between students' learning, alumni evaluations, improvement of courses and instruction, and recognition of excellent teaching performance (Wachtel, 1998). Richardson (2005) reported that the student evaluation instrument is an effective tool that provides important feedback in regards to effective teaching and the need to improve the quality of courses. However, there is no documentation in the literature that the feedback is used for this purpose; nor is there documentation of a standard, strategic follow-up in the annual review process to assess faculty attention. The sole means of assuming faculty attention is typically attention to subsequent course evaluations.

On the other hand, Hornstein (2017) concluded that student evaluations should not be used as a criterion to measure teaching performance because of potentially low validity, response rates, students' biases, and the pressure for faculty to make the course easier for their students, so that they may receive a higher evaluation. The common problem of student evaluations that defines whether faculty finds it acceptable or

misrepresented is the ratings that are received. Those who receive lower course evaluations tend to believe course evaluations are misused and invalid; and those who receive higher course evaluations accept the tool as valid. These beliefs are based on faculty self-perception of their teaching effectiveness after reviewing student evaluations which affect their attitude towards this instrument and if the tool is reliable during the evaluation process for promotion and tenure (Roche & Marsh, 2002).

Though students are a great source of providing feedback in this domain due to majority of time spent within the course and numerous interactions with the faculty member, they may not be as knowledgeable of the standards of effective teaching or aware of the behavioral indicators of instructors meeting the requirements. The importance of effective teaching in higher education from the faculty viewpoint may differ from those of most students, causing some faculty to question whether this is a fair measurement of their teaching effectiveness (Feldman, 1988). This difference in perspective may affect the reliability and validity of faculty overall course evaluation score. Furthermore, the controversy about the use of a long-standing measure of effectiveness in the domain of teaching by an untrained student evaluator has life-long implications in the career of university-level instructors including: annual evaluations that influence salary increases; promotion, which is attached to salary increases, career advancement, and continued employment; and tenure, which is attached to job security and longevity (Stark-Wroblewski, Ahlering, & Brill, 2007).

Summary and Statement of the Problem

        Currently university course evaluations are noted to be the most prevalently used instrument to measure the quality of teaching effectiveness (Haladyna & Amrein-Beardsley, 2009; Wright & Jenkins-Guarnieri, 2012). Previous research has shown this tool to be quite popular in providing sufficient feedback to improve the learning environment and instructor presentation. Course evaluations are also useful for the supervisors and committee members who review faculty dossiers for progress in teaching performance, service, scholarships, and publications or presentations to meet criteria for promotion and tenure. The course evaluations are taken in to consideration during this annual review to gain students perspective of the faculty member effectiveness as an instructor and advisor; the availability of the information provided by these individuals also benefits future students' knowledge of the instructor and the course before enrollment (Chen & Hoshower, 2003). Such information is also considered in determining merit, which is associated with increased compensation.

        While some educators believe course evaluations are beneficial, there are some who will dispute otherwise due to the factors that has a negative influence on their rating as an instructor. Faculty and staff disagreement rise from concerns of whether a student is qualified to be a rater with this much power to influence faculty careers, and if their perspective is a true valid and reliable measurement of effective teaching (Darwin, 2012; Emery, et. al, 2003). Students' rating may be influenced by items such as their grade, interest in course, attractiveness to faculty member, and personal expectation of the

faculty member which can cause them to be bias and rate instructors evaluation with a

subjective view (Nargundkar & Shrikhande, 2014). Student evaluations of teaching

effectiveness are also affected by response rates due to lack of time or the belief that their

opinion does not matter. Often this is the complaint of some students who choose not to

participate in the evaluation of instructors: "It won't make a difference anyway?" This is

often the outcome when students dislike an instructor or may be fearful of sharing their

commentary (Winchester & Winchester, 2012). These potential issues cause some

educators to question whether the course evaluation is focus more on popularity than

educational concerns.

Course evaluations are found to be a common measurement used across

universities because it is efficient measurement that can be used to obtain valuable

information to document the quality of the learning environment, while also making the

necessary changes to enhance the program (Spooren et. al, 2013). With student

evaluations being a necessary measurement to collect for teaching effectiveness and

promotion and tenure decisions, it is imperative to consider the advantages and

disadvantage of the influence it has on faculty overall course evaluation. The focus of this

study is to examine the following research question:

1. To what degree do faculty demographics (i.e., race, gender), academic

    background (i.e., years since completion of terminal degree?, status of institution

    where degree was awarded), academic status (i.e., terminal degree, tenure vs. non-

    tenured, faculty rank, faculty work/class load), professional engagement as noted

on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in overall course evaluation ratings?

2. To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree, status of institution where degree was awarded), academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in the standard deviation of the course evaluation?

3. To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree, status of institution where degree was awarded), academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in response rates?

The second question was included to examine the contribution of all independent variables on the criterion variables: standard deviations of the course evaluation. A correlation matrix examining the relationship among all pairs of variables will allow a check for collinearity. This review process will determine the final variables to be included in the proposed regressions. The third question was included because the seldom addressed variable of students' response rates was considered key in determining the validity of the cumulative course evaluation. Each variable will be explored to identify how much influence it has when students evaluate their courses and identify the strongest predictor in the model for overall course evaluation rating. A comprehensive literature review of prior research that examined variables that influence course evaluation will be presented and discussed in Chapter 2.

Chapter II

Literature Review

As most universities are using course evaluation as the only mean of measuring teaching effectiveness, minor research has been conducted to examine factors that influence student rating of teaching and the impact it has on tenure and promotion. The purpose of this study is to examine how faculty demographic, academic degree, academic status, professional engagement, grading patterns, and course variables as independent variables predicts or significantly impact the dependent variable, the university faculty cumulative course evaluation, standard deviation of course evaluation, and response rate. The relationship between each of these variables is hypothesized to influence student participation and motivation when evaluating the quality of their courses and faculty member teaching performance (Beran & Violato, 2005; Isely & Singh, 2005; Kogan, Schoenfeld-Tacher, & Hellyer, 2010; MacNell, Driscoll, & Hunt, 2015; Marsh & Hattie, 2002; Nulty, 2008). The literature review that follows provides an overview of the prior empirical studies and conceptual articles associated with the study of university course evaluation.

Faculty Demographics

Based on prior research, faculty demographic is found to have quite an impact on student evaluations of faculty teaching effectiveness and the promotion and tenure process (Basow, 2000; Bavishi, Madera, & Hebel, 2010; Ho, Thomsen, & Sidanius,

2009; Sinclair & Kunda, 2000; Sprague & Massoni, 2005). The gender and race of a faculty member have been found to be associated with students' subjective view when rating faculty member teaching performance based on their bias perceptions. Reid (2010) examined the influence of instructors' race and regards to the promotion and tenure process which may hinder faculty of color to prosper in their careers to the same degree as White faculty members. It further suggests the importance of acknowledging how race-based biased may affect reliability and validity in students' evaluation of their faculty member teaching performance.

Because of the design of this study, some may quickly assume that the findings raise questions about the overall varied degree of competence in teaching based on racial-group status in predominately White classrooms. However, although the means were significantly lower, within-group variance in the course evaluation ratings among each racial/ethnic group of faculty did exist. Some students rated racial ethnic minority faculty higher than others, just not sufficiently to effect overall cumulative rating. It appears that it may be this variance that lowers the overall rating in comparison to White peers. Even if the critical number of students rates a faculty member favorably, the ratings of a few, who may be influenced by bias and race-based varied expectations, will lower the overall evaluation. Subsequent study of this phenomenon is certainly warranted.

In a more controlled study, MacNell, Driscoll, and Hunt (2015) conducted an experiment to view the influence of gender bias on students' evaluation of faculty members teaching effectiveness. In this controlled experiment, they created an online

class with assistant instructors who taught one class with their own identity, and taught the other class under a different gender identity than their actual gender. For example, Group A instructor's perceived gender and actual gender was female; Group B instructor's perceived gender was female and actual gender was male, Group C instructor's perceived gender was male and the actual gender was female; and Group D instructor's perceived gender and actual gender was male. Findings led the researchers to conclude that gender bias did exist. Students rated the perceived male teacher evaluation significantly higher than the female instructor despite the instructor's gender identity. Specifically, perceived male received higher scores than perceived female, actual male, and perceived female on professionalism, respectfulness, praise, and promptness questions. Despite the similar teaching strategies and interactions during online class discussion, inequality was still shown among the students as they rated the female faculty member more harshly than her male counterpart. The gender expectations that exist for both male and female may influence students to believe that male teachers should be masculine and hold higher status jobs, and deserving of greater respect, while female teachers should be more feminine, supportive, and warm; if they are not as anticipated, course evaluations are negatively affected. When these expectations are not met students assume female teachers are not competent in their position and evaluate these faculty members lower than they would a male faculty member who does or does not meet their gender expectation (MacNell et al., 2015). This experiment brings much insight about the expected gender roles and how it may influence students' evaluation of their

experiences, this point indicates that course evaluations may be affected by attitudes related to gender that students bring to the classroom, and not what happens in the classroom.

If female faculty are expected to be warmer, accessible, and more supportive than male colleagues, then they may be rated lower for exhibiting the exact same engagement behaviors. For example, Sinclair and Kunda (2000) conducted a study that focused on positive or negative feedback (grades received) provided by the instructors and how it influenced students' evaluation of their male and female instructor. Based on their finding, the instructors who provided students with a higher grade were rated more positively than instructors that provided low grades. The researchers found that students who received a lower grade from a female instructor rated her less competent than the male instructor and female instructor who provided a higher grade. It appears that students differ in expectations of female and male instructor, therefore this stereotype of female instructor influence student evaluations to rate their instructor poorly when expectations are not met. These are outcomes that faculty cannot control without penalty to the faculty and the students' learning and exposure to rigor. Consequently, female faculty may be forced to over-accommodate and over-extend themselves (work harder) just to receive the same outcomes that male colleagues receive with less work and engagement. For example, accessibility is an item included in some course evaluation measures. This gender difference might be evident in students' expectation that female faculty should be more accessible than male faculty. These findings suggest that this

expectation would result in a lower rating on the item for the female faculty member even if the degree of accessibility was the same as with her male colleagues.

Further investigation of gender bias is explored when Centra and Gaubatz (2000) examined the difference in male and female student evaluations for both male and female instructors. The focus of this experiment was to compare evaluation rating of both faculty members from female and male students within the same classes. Also, all male ratings and then all female ratings were used to determine the difference between male and female instructors (Centra & Gaubatz, 2000). The researchers used a student evaluation form consisting of seven scales including 1) course organization and planning, 2) communication, 3) faculty/student interaction, 4) assignments, exams, and grading, 5) course outcomes, 6) student effort and involvement, and 7) course difficulty, workload, and pace. Based on their results, female students rated female instructors higher across six of the eight scales while male instructors received equal rating from all students. Their study suggests gender preferences in which female students were more likely to enjoy the teaching methods of a female faculty member than a male faculty member who focused- more on lecture than discussion (Centra & Gaubatz, 2000). Researchers concluded that gender bias does exist in faculty members' course evaluations, but it does not determine whether students learn more or less based on teacher effectiveness across the courses. Being aware of bias is important when reviewing course evaluations, because of the various factors that may influence a student rating whether it is gender bias,

motivation, interest, student-faculty member interaction, expectation, or class size that may affect the validity and reliability of an instructors' evaluation.

Faculty Academic Background

The university structure is based on the assumption that doctoral level faculty are more knowledgeable and experienced than masters level faculty. This notion implies that the faculty years in terminal degree and status of institution where degree was awarded may influence higher students' evaluation of faculty' teaching effectiveness. Yet, the effects of this particular faculty demographic on course evaluations have been rarely examined. In one study on the association between instructors' demographic information and students' ratings of instructional effectiveness, Johnson, Narayanan, and Sawaya (2013) examined the instructor's experience and academic rank as factors that affect teaching performance. Contrary to their hypothesis, their results were that instructors with more experience and who was identified as tenured had significantly lower rating on their course evaluation compared to the non-tenured faculty.

However, in contrast, previous research such as McPherson, Jewell, and Kim (2009) show that tenured faculty received higher evaluations than non-tenured faculty; and support the notion that with experience, faculty are more likely to receive lower evaluations possibly due to less preparation and interaction after years of teaching the same class. Yet, these results differ from Kinney and Smith (1992) findings, in which they report that more positive student evaluations come with faculty who are older and more experienced. Research has shown mixed results and indicates that there is a

relationship between experience and student evaluations, but the direction of this relationship varies. Though faculty may have years of experience in teaching a specific course and have obtained tenure, student evaluations can still be affected by other characteristics such as productivity, age, or enthusiasm (Clayson, 1999; Stonebraker & Stone, 2015; Tang & Chamberlain, 2003). Some students may give a lower evaluation for a course because they prefer being taught by a younger and more enthusiastic faculty member. Often times, adjunct faculty possess these specific characteristics during their entry into the field. The influence of faculty-specific variables has been examined, but in isolation and with starkly different findings.

The sole purpose of hiring an adjunct faculty is for teaching. Without the requirement of attending to additional activities associated with promotion and tenure (i.e., research/scholarship, professional service), adjunct faculty can concentrate more on preparation for instructions and instruction, which intuitively should result in higher evaluations. There is no empirical support for this hypothesis in the literature. Wollert and West (2000) found that adjunct faculty members were rated higher than full time faculty in most of the subscales of the students' ratings. The five subscales used to make up the questions on the evaluation included: attitude, methods, content, interest, and instructor. Specifically, the adjunct faculty and instructors scored higher on the interest, content, and method subscales. In many universities, adjunct faculty are not required to partake in research or scholarships, therefore this may allow them time to concentrate on teaching which appears evident in this study. Without the requirement of research and

scholarship, this may put more pressure on adjunct faculty to focus on receiving higher evaluations and reinforce some reluctance to give lower grades, even when deserved based on established criteria (Sonner, 2000).

Previous research has shown that there is difference between experience and responsibility between ranks that is linked to student evaluations of teacher effectiveness. Despite the minimum experience, low resources on campus, and increase in class load, adjunct faculty evaluations are still higher or equivalent to full-time faculty evaluations (Landrum, 2009). This point indicates that course evaluations are affected by teaching capability, and not entirely on the number of years dedicated to teaching the core curriculum as a full-time or adjunct faculty. There are few studies that indicate faculty members' years in terminal degree and status of institution where degree was awarded influence student evaluations. It is possible that these variables do not influence student perceptions when rating faculty member.

Faculty Academic Status

According to Mohan (2011), employing numerous non-tenure track faculty may influence student evaluations and grade distributions. Students' ratings have been found to be significantly associated with the faculty members' academic status. In this study, non-tenured faculty and tenured faculty in business schools were investigated to determine how student evaluations and grade distributions differ between the faculty members and the impact they have on tenure and promotion. The results of this study

found that non-tenure-track faculty to be rated higher in student evaluations than the tenure track faculty.

Isley and Singh (2007) found support that higher grade expectations led to better student evaluations, but differs across faculty rank. Adjunct faculty course evaluations were found to be significantly more affected by students' grade expectations. Due to these findings, the question that is created is whether student evaluation of teaching (SET) is measuring the faculty teaching performance or just student satisfaction with their anticipated final grade within the course (Mohan, 2011). It can also be assumed that students may perceive non-tenured faculty as better teachers. The required engagement in scholarships and professional services for those on tenure track may be a distraction from optimal levels of preparation for instruction and mentoring of students. Though some institutions of higher education may find it beneficial to have a mixture of non-tenured faculty and tenured faculty to teach courses at a lower cost, it could be harmful when the focus of a non-tenure track faculty is solely on receiving high student evaluations. This motive may lead to change in the course content with lower expectations, easier exams, and reduction in workload just to please students and secure higher course evaluations for job security (Mohan, 2011). While tenure-track faculty have three primary domains on which to base job security (i.e., teaching, research, and service), adjunct and non-tenure-track faculty have only one sole focus. Tenure track faculty may, be stretched to provide the optimal degree of special attention that students might prefer in course instruction because of required commitment to scholarship and service; consequently, they may not

be as accessible as non-tenure track faculty, and receive lower course evaluations. On the other hand, non-tenure-track faculty may have to be monitored and supervised to insure coverage of course content and effective delivery, which meet program standards.

Liu (2012) examined the factors that influence student evaluation of instructors in online classes including gender, class size, reasons for taking course, student class status, and instructor's rank to measure the quality of teaching across 29 colleges and universities. The student evaluations for the course were collected using an online program which consisted of questions pertaining to 1) course organization and planning, 2) communication, 3) faculty/student interaction, 4) assignments, exams, and grading, 5) course outcomes, 6) student effort and involvement, and 7) course difficulty, workload, and pace on a two-level hierarchical model. Descriptive statistics and confirmatory factor analysis was also used during this study. Results indicated gender and class size did not affect student evaluations as significantly as the other factors which greatly influenced student evaluations (Liu, 2012). Required courses were rated lower than elective courses; upper-level classes were rated higher than lower-level classes; and faculty rank was a significant predictor of student evaluations in which assistant and full faculty members were rated lower than other instructors. Liu proposed that the results may be due to students having an option to attend course based on interest level, difference in college experience between graduate and undergraduate student, and difference between faculty workload and requirements based on their academic status (Liu, 2012). While tenure-track faculty have a mandate to focus on research and publication (e.g., "publish or

perish"), non- tenure track faculty are more likely to focus on being a more effective teacher which may explain the difference in student evaluations.

Nargundkar and Shrikhande (2014) also found that ratings were higher for non-tenure track faculty than tenured track faculty on student evaluations. Yet, tenured faculty ratings were high for teaching graduate courses. Faculty academic status in regards to being an adjunct, clinical, assistant, associate, or full-time faculty member may influence the overall scores of course evaluation due to work/class load, research, and level of teaching priority.

Faculty Professional Engagement

It is assumed that faculty who are the most current in the literature and most visible in leadership activity within professional organizations are also the most knowledgeable of cutting edge information. This is one component of being an 'academician' and insuring successful upward mobility on the tenure-track. The controversy about the relationship between faculty scholarships, productivity, and teaching effectiveness remains up for debate. Some examples of this dispute between academicians comes from a study by Prince, Felder, and Brent (2007) who examined whether research productivity influences teaching in undergraduate courses. They report that most faculty and administrators argue that research does improve teaching, some see the possibility of research being an influence, and while others have reported an insignificant relationship between teaching and research productivity (Gray, 1992; Marsh & Hattie, 2002). In one study, Allen (1996) found a small positive significant relationship

between teaching effectiveness and engagement in research after gathering literature review that examined teaching effectiveness, research productivity, and statistical information of effect size. Specifically, variables such as student evaluations, peer evaluations, teaching awards, and amount of teaching related activities demonstrated a positive correlation between teaching and research productivity when used to evaluate the quality of instruction. The number of publications, teachings awards, and peer ratings were slightly higher than research productivity. While research is a criterion that universities require faculty to partake in for tenure and promotion, there are studies that have shown research to not be as beneficial in regards to teaching effectiveness. For instance, Hattie and Marsh (1996) found no significant relationship between research productivity and teaching effectiveness after examining studies that explore multiple variables including self-ratings, publications, and faculty satisfaction with involvement in teaching and research. Literature reviews are limited on the relationship between engagement in research and student evaluations; studies that have been conducted does not necessarily prove whether teaching effectiveness and research productivity enhances each other (Centra, 1983). If research is important to teaching effectiveness as most academicians purport, questions are raised about the reason research is not a requirement for adjunct faculty. Also, it questions the structure of universities by having majority of non-tenured faculty teach undergraduate courses and tenured faculty teach mainly graduate courses (Barnett, 1992). Furthermore, it can be assumed that research participation, interest, and resource are limited to undergraduate students compared to

31

graduate students, compared with graduate students, who are encouraged to partake in research as they prepare for their field of study. Based on this literature review, it would be hypothesized that course evaluation ratings would be negatively correlated with scholarship productivity. The greater productivity in scholarship, the lower the course evaluations.

Faculty Grading Patterns

In a study of course evaluations, Brockx, Spooren, and Mortelmans (2011) examined the influence of grade distribution, course content, faculty, and student characteristics on student evaluations of teaching (SET). Brockx et al. (2011) found that course evaluations were found to be positively correlated with high course grades, class attendance, and examination scores. Class attendance is one of the student characteristics found to impact SET more than student gender. Students who attend class regularly are likely to give higher evaluations because of the significant amount of time spent in the classroom and exposure to their faculty member teaching performance. Elective courses, field of study, and level of the course are a few course factors found to significantly influence SET. For elective courses, students are more motivated about going to a class they were able to choose versus a required course which could influence how they evaluate the faculty member. For the field of study, areas of art and humanities were found to have a higher rating of evaluations than the field of biology or math (Brockx et. al, 2011). The level of the course also impacted the SET, for example faculty members with easy or undergraduate classes may have higher student evaluations than the graduate

and challenging courses other faculty members teach. While evaluating characteristics of faculty members, researchers found the teacher's age had minimal influence while teacher's rank showed inconsistent results. The relationship between grade distribution and SETs has shown faculty members who give higher grades can expect to have higher evaluations than faculty members who give lower grades to their students. Each of these factors was found to be significantly correlated with the rating students give their faculty member teaching performance.

Previous studies have also shown similar results of this relationship, which found course grade to be a significant predictor of SET (Krautmann & Sander, 1999; McPherson, 2006). If students rely heavily on grades while rating evaluations, it can be assumed that this measurement may not be as valid and reliable and contain possible bias due to dislike of the faculty member and dissatisfaction with final grade. Also, with student evaluation being a measurement to improve teaching and for promotion and tenure, this may place higher pressure on faculty to give higher grades to obtain a better evaluation and affect the credibility of SET. However, Heckert, Latier, Ringwald, and Silvey (2006) believed that the relationship between course grade and SET is a positive indication of teaching effectiveness. They report that students who receive higher grades from faculty members, who encourage learning, and teach effectively result in a better evaluation. Despite the grade that was received, they also consider other variables including the course, instructor, and student characteristics that could have also influence the low and high ratings of the SET. In accordance with grades received, students

evaluate how much they learned from the class, the workload given, and the difficulty or ease of the course. For example, students gave low ratings on evaluations of courses that were too difficult or easy, and provided higher ratings for those that were fair because they learned more (Centra, 2003). When examining the influence of factors that impact overall course evaluations, the relationship between students' grades and course evaluation is predicted to be a significant factor in satisfactory and unsatisfactory evaluation scores from previous research. Langbein (2008) concluded that faculty members who give higher grades result in a satisfactory SET and higher pay. Grade inflation decreases the quality of education and places more focus on satisfying the student as consumer and the faculty members desire to have a better chance at promotion and tenure. Based on findings from the literature, it may be hypothesized that students' anticipated course grade and faculty members' tendency to have a greater frequency of 'A's will have a positive and significant relationship with overall course evaluations. Students who anticipate higher grades will tend to evaluate course instruction more positively; faculty who tend to assign higher grades will tend to receive higher course evaluations.

Course Variables

While variables such as race, gender, academic status, and grading patterns have been explored to a degree of impact on course evaluation, research has also been conducted to provide understanding of the relationships between course evaluations and variables associated with descriptors specifically related to the courses. For example,

34

course delivery and course levels may also significantly impact students' rating of instructors and instruction. According to Young and Duncan (2014) when comparing online to face-to-face classes the SET scores differ based on students' perception of the teaching environment. As online student evaluations are becoming more efficient, online classes are in high demand as it becomes more convenient for students and faculty (Allen & Seaman, 2013; Chau, 2010). Based on their research, results did indicate that face-to-face classes were rated higher than online classes in communication, faculty member and student interaction, grading, teaching methods, course outcome, and overall evaluation, while student effort was rated higher in online classes (Young & Duncan, 2014).

Likewise, in previous studies, evaluations were found to be significantly higher in face-to-face classes than online classes because students felt more motivated and had greater satisfaction with their faculty member being present in the classroom setting (Mentzer, Cryan, & Teclehaimanot, 2007; Summers, Waigandt, & Whittaker, 2005). The findings may indicate that students expect the same teaching effectiveness of faculty members in online classes as they receive in face-to-face courses. When the criteria are not met in areas such as faculty interaction, enthusiasm, and communication, students may tend to become dissatisfied with the course and provide lower ratings on evaluations. The issue that then arises is the amount of control faculty has to provide these accommodations to meet students' expectations in online courses. However, studies have shown that with the combination of traditional and online classes, hybrid courses offer more benefits for students to receive quality instruction and participate in more

interaction with their faculty member and peers (Riffell & Sibley, 2005; Tuckman, 2002).

Findings from these studies have shown that students rated hybrid courses slightly higher

or equivalent to traditional courses and higher than online courses. The hybrid course

offers students the ability to attend class and engage in interaction with their faculty

member, while also having the option to self-pace in learning activities and assignments

on the web. When examining the significant difference among variables in each

environment, it can be argued that students in online classes lack the interaction,

engagement, and motivation that are adequately provided in a hybrid and face-to-face

course.

Moreover, there have been very few studies of the relationship between student

evaluations of teaching effectiveness and the course level, but those that have been

conducted have shown mixed results. For example, Young and Bruce (2011) found that

students' evaluations of instructors, who taught upper level courses were rated higher

than instructors who taught the lower level courses. Whereas, studies such as Marsh and

Overall (1981) have shown the level of the course is relatively less important in

determining the outcome of student ratings than the instructor who teaches it. After

examining the influence of course level, course type, and assigned instructor on SET,

their findings concluded that the course level and course type was significantly lower

than the instructor assigned to the course.

Beran and Violato (2005) further examined how specific student and course

characteristics may influence student ratings on course evaluations. In their study, the

variables considered were the student expected grade, class attendance, workload, program, course status, type of course, and term of course. While reviewing these variables, the researchers examined how courses such as lecture, lab, and practicum-type courses contributed to student evaluation of teaching. Ratings differed significantly for students, who had lab courses, reported higher evaluations than those in lecture classes (Beran & Violato, 2005). Information provided from these results may suggest that students who take lab and practicum classes feel a sense of accomplishment when completing task independently which results in higher student evaluations than those enrolled in lecture status courses.

Overall Course Evaluation and Response Rate

The response rate is an important dependent variable to review because of the significant influence it has on the overall course evaluation (Fosnacht, Sarraf, Howe, & Peck, 2017). This variable is based on the number of students that participate in providing feedback about the course content and faculty member teaching effectiveness within the class, which then factors into the overall validity of the course evaluation. Previous studies have examined variables that might be significantly associated with students' response rates in the evaluation of university courses and instructors such as access to technology, benefits of participation, extra-credit and other incentives, e-mail reminders, and remembering to participate (Anderson, Brown, & Spaeth, 2006). For example, Guder and Malliaris (2013) studied factors associated with undergraduate and graduate students' motivation to complete online course evaluations and the influence it has on response

rates. They hypothesized that students would be more eager to give their opinions about their course experience with the idea of expressing themselves freely and using the online system to provide their feedback. Based on their research, they found that response rates decreased when the in-class survey was switched to online surveys (Guder & Malliaris, 2013). Though this significant decrease occurred, universities still rely on giving online surveys for students to access and provide feedback for course evaluations. To determine why this occurred, the researchers explored four possible reasons: students being disengaged, technology problems, perception of no benefit from participation, or other reasons (Guder & Malliaris, 2013). A questionnaire was given to students to collect information about whether they received an initial and reminder email as well as encouragement from the faculty member about completing the evaluation. Also, the researchers collected information about whether the students completed all, some, or none of their course evaluations. Response rates increased when graduate and undergraduate students were reminded by email to complete the online course evaluation and also when their faculty member took a moment to inform the students the importance of completing course evaluation during class time. Further information about students' interest in the course, course load, and incentives to complete the course evaluation could also been used as a measure to identify what motivate students to complete the course evaluation (Guder & Malliaris, 2013).

Hatfield and Coyle (2013) examined the relationship between course evaluations and the following: student's grades, gender, age, and ethnicity. Students' participation in

the course evaluation process was found to be significantly related to gender, age, and ethnicity more so than anticipated grades receives. Older, Caucasian or Asian female, students tended to participate on course evaluation more frequently. Similarly, Schiekirka & Raupach (2015) found female students to be more likely to give positive ratings for course evaluations. These findings suggest that student' characteristics influence participation in engaging in providing feedback on course evaluations.

Also, it could be argued whether students' interest and perception of course evaluations influence their desire in engaging in evaluations (Marlin, 1987; Gaillard, Mitchell, & Kavota, 2011; Spencer & Schmelkin, 2002). It is probably those who do not view evaluations as valuable and tend to believe that their evaluations are unimportant and don't have an effect on courses or teacher performance, that do not participate in the evaluation process.

The response rates affect the validity and reliability of the overall course evaluation in several ways. For instance, when examining both the response rate and overall course evaluation the outcome may result in 1 of 4 ways: 1) high response rate and high overall course evaluation, 2) high response rate and low overall course evaluation, 3) low response rate and high overall course evaluation, or 4) low response rate and low overall course evaluation (Al Kuwaiti, AlQuraan, & Subbarayalu, 2016). Some might hypothesize that low response rates might tend to be associated with high overall course evaluations because the students who were dissatisfied with the faculty member and course may choose not to respond. On the other hand, low response rates

might tend to be associated with low overall evaluations because students who are dissatisfied with course are more likely to respond than students who are satisfied (Simpson & Siguaw, 2000; Williams & Ceci, 1997). It is equal opportunity for both students that do well or poorly to participate in course evaluations. Neither pattern is considered in reviews for promotion and tenure or annual evaluations of teacher effectiveness.

Nulty (2008) examined the response rates between online and in-class course evaluation surveys and provided suggestions to determine when a response rate is large enough to be adequate. He found that the online survey response rate decreased by about 23% in comparison to the in-class response rate from student teacher evaluations (Nulty, 2008). The lower the response rate the less valid the overall course evaluation will be. For example, if a faculty member has a perfect score for their overall course evaluation but only 4 students provided feedback out of a class of 20 than the course evaluation would be considered invalid. Without sufficient information from other students within that class, it would be difficult to determine the faculty member true teaching effectiveness. Nulty (2008) explains that the low response rate is not a good representative of the students within the course, which causes bias in the overall course evaluation score. A specific percentage of response rates should be required to obtain a valid score of course evaluations before it is included in the reviewing for faculty member tenure and promotion process. Some research articles suggested 50% to 70% to be an attainable and desirable for an adequate response rate for course evaluations

(Baruch, 1999; Richardson, 2005). An institutional-designated percentage for response rates provides an average requirement of student participation for course evaluations within a given class as a justification of knowing what is acceptable and valid when reviewing faculty's dossier. When evaluating the overall course evaluation, it is important for the supervisor and committee members to identify and examine the percentage of both the response rate and overall course evaluation for validity. In the literature addressing student course evaluations, there is currently no study that empirically examines all the variables at one time. This study will add to the literature by providing a more in-depth examination of the meaning of course evaluations.

Course evaluations are being used in the annual review process to assess the competence of instructors for promotion and secure employment; while also being used as a tool to improve the quality of learning in courses. The specific purpose of this research is to examine variables associated with the variance of course evaluations. In particular to determine:

To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree?, status of institution where degree was awarded), faculty academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), faculty professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course

level, instructional delivery type, and course type) predict the variance in overall course evaluation ratings, standard deviation of the course evaluation, and response rates. It is hypothesized that faculty characteristics, grading practices, and course types would be significantly associated with students' rating of courses and students' response rates.

Chapter III

Methods

Procedure

This study was conducted from archival data from an academic unit housed in a college of education. Securing IRB approval and permission to access the course evaluation data was obtained before gathering faculty information from this academic unit. The university website was used to collect instructors' vita using any secure technology device that has internet access. Individuals' identity will be protected by coding and through maintaining collected campus-wide data, instead of identifying results by program or department. To ensure confidentiality of the data, faculty names and any other identifiable information was removed. Each instructor was assigned a specific code number corresponding to the data collected for each individual. The codes were stored in a locked filed and downloaded to one computer that is password protected within campus grounds.

Given that archival data was used, there will be no need for informed consent. There was no compensation involved within this study. The risks were minimal given that most of the demographic and faculty vitae content included in this study were archival data, which is available to the public. Once copies of teaching evaluations, annual evaluations, and faculty vitae were printed, materials were secured in a locked cabinet in the principal investigator's office, which was always locked and unsecured.

The results will provide a positive contribution to the literature in higher education and will increase understanding of the meaning of course evaluations in the review process for annual evaluations, hiring, and promotion and tenure. The information will be beneficial to students who aspire to become faculty, tenure-track faculty, and administrators who establish policy related to salary increases, hiring, and tenure and promotion.

The data analyzed was based on the archival data of a total of fifty-five ($n= 55$) faculty members that were classified according to academic rank: 26 (48%) adjunct, 8 (14%) clinical faculty, 5 (10%) assistant, 8 (14%) associate faculty members, and 8 (14%) full faculty members within the department. Within this sample of faculty members the following information was gathered: 32 (58%) with a master degree and 23 (42%) with a doctorate degree, 34 (62%) non-tenure and 21 (38%) tenure, 29 (53%) obtained terminal degree from a non-research institution and 26 (47%) obtained terminal degree from a research institution, 21 (38%) considered as part-time faculty and 34 (62%) considered as full-time faculty, 38 (69%) females and 17 (31%) males, and 50 (90%) faculty members who identify as racial major and 5 (10%) faculty members who identify as racial minority. Each instructor's course load assignment varied and course evaluations were collected from a total of 284 courses department wide. The department also includes two courses designated as Core Curriculum, which guarantees the inclusion of student majors not housed within the department.

Dependent Variables

In addition to the background information described earlier, the course evaluation contains questions about both course content and instructor effectiveness for the use of course and program improvement and for instructor evaluation. The course evaluation used in this study represents an example of a traditional measure. The college of education course evaluation rating is on a 5-item scale ranging from 5 = very good, 4 = good, 3 = average, 2 = poor, to 1 = very poor (see Appendix for college of education course evaluation).   A column is also provided for students to select no opinion and an option to provide additional comments. The dependent variables being influenced in faculty course evaluations include: *cumulative course evaluation* – the overall average evaluation score of the course, or the average score of students' responses; *standard deviation of the course evaluation* – mean of faculty cumulative course evaluation; and *response rate* – the percentage of students in a course responding to the evaluations, or the number of students who completed the course evaluation divided by the number of people in the course.

Independent Variables

Variables manipulating the variance in faculty course evaluations include: *faculty demographics* – the ethnicity and sex of faculty members (racial/ethnic minority vs. racial/ethnic majority, female vs. male.); *faculty academic background* – the year of their terminal degree and if the degree was awarded from a research vs. non-research institution; *faculty academic status* – the rank of faculty members who are tenure-track

vs non-tenure track, masters-level vs. doctoral level, and adjunct, clinical, assistant, associate, and full-time faculty members, and part-time vs. full-time; *faculty professional engagement* – the number of presentations, publications, and professional license/certifications that was gathered from faculty members updated curriculum vitae during the academic year course evaluations were collected; *course grading patterns* – the number of As, Bs, Cs, Ds, and Fs assigned during the academic year course evaluations were collected; and *course characteristics* – the instructional delivery type (online vs. hybrid vs. face to face), course type (lecture vs. practicum), and course level (undergraduate vs. masters vs. doctoral).

Analysis

Pearson Product Correlations will be examined to initially identify any significant relationships that may approach the level of multicollinearity (Cho, 2017; Kraha, Turner, Nimon, Zientek, & Henson, 2012; Martz, 2013). These results will guide decision-making about the final selection of independent variables to be included in the regression model. Due to the exploratory status, a backward step-wise regression analysis will be used with the dependent variables to identify the best model. The backward multiple regression analysis will be used to identify the association between two or more variables, specifically the relationship between a set of independent variables and a selected dependent variable (Wilkinson, 1979). When conducting research that involves testing multiple independent variables to predict one dependent variable, a backward regression equation is recommended as a common statistic test that best enter and remove

these predictors (Marsh, 1980; Thomas & Galambos, 2004). This approach brings attention to various influential factors in relation to the dependent variable. Backward regression will be used for this study because the research question contains multiple predictor variables that might possibly influence the criterion variable.

For the purpose of this study, a backward regression tool is used to identify to what degree faculty demographics, faculty academic background, faculty academic status, faculty professional engagement, course grading patterns, and course characteristics predict the variance in overall course evaluations, standard deviation of the course evaluation, and response rates. These predictors were selected based on previous research conducted to determine its relation to the criterion variable (Brockx et al., 2011; Centra, 2003; Heckert et al., 2006; Marsh & Overall, 1981; Nowell, Gale, & Handley, 2010). By using the regression strategy, an equation will be developed to combine these predictors. The backward regression analysis will be an effective tool to determine the contribution of each independent variable to the dependent variable and overall variance. This is a non-hypothesis testing analysis.

Using the information collected from faculty vitae and the course WebFocus site, in a larger regression model with faculty cumulative overall course evaluation, standard deviation of the course evaluation, and response rate as the dependent (criterion) variable, and faculty demographics, academic background, academic status, professional engagement, course grading patterns, and course characteristics will be included as

47

independent variables in a regression equation. The variables incorporated in the model were the following:

Instructor-related variables will include faculty members' personal demographics (i.e., gender, race/ethnicity status), professional related information including faculty members' academic background (i.e., year of terminal degree, status of institution where degree was awarded), faculty members' academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), and faculty members' professional engagement (i.e., # presentations, # publications, # licenses/certifications). Coding will be assigned for each of the independent variables as follows:

Gender: coded: 1 = female, 2 = male

Ethnicity: coded: 1 = racial majority, 2 = racial minority

Institution where degree was awarded: coded: 1 = non-research, 2 = research institution

Tenure: coded: 1 = non-tenure, 2 = tenure

Terminal degree: coded: 1 = MA, 2 = PhD

Faculty rank: coded: 1 = adjunct, 2 = clinical/instructor, 3 = assistant, 4 = associate, 5 = full

Faculty work/class load: coded: 1 = part-time 2 = full-time

Instructor course-related variables included course grading patterns (i.e., number of As, Bs, Cs, Ds, and Fs assigned), and course characteristics (i.e., course level: undergraduate vs. masters vs. doctoral course levels, instructional delivery type: online vs. hybrid vs. face to face, and course type: lecture vs. practicum).

Instructional delivery type: coded: 1 = none online (face-to-face), 2 = some online

(hybrid), 3 = all online

Course type: coded: 1 = lecture, 2 = practicum

Course level: coded: 1 = undergraduate (100-400 level), 2 = master (500 level), 3 =

doctorate (600 level)

Administration-related variables will include the number of courses each faculty

assumed during the semester of the course evaluation. Independent studies, theses, and

dissertations were not included in the study.

Chapter IV
Results

University Setting

The Institutional Review Board (IRB)-approved study was based on the analysis of archival data consisting of faculty members' course evaluations during one academic year, in an on-campus department with 8 majors and a student enrollment of 572 during the time of data collection. There was no participant recruitment process; archival data was used. Data were obtained from a BANNER system, an electronic data interchange system from a four-year teaching and research university. The state university is a comprehensive, regional institution which enrolls approximately 13,000 students and offers approximately 80 undergraduate majors and more than 120 areas of study and concentrations within six academic colleges including business, education, fine arts, forestry and agriculture, sciences and mathematics, and liberal and applied arts. The university offers degree programs both online and on campus. During the school year, the institution of higher education had a student body of approximately 12,999 and a faculty body of approximately 747. The student-to-faculty ratio is 20:1 and an average class size of 29.

Description of College. The college of education offers 27 undergraduate/graduate degree programs and 45 educator certifications. It consists of five departments including: Elementary Education Department, Human Services Department,

Kinesiology & Health Science Department, Secondary Education & Educational Leadership Department, and the School of Human Sciences. The college embodies access, equity, diversity, cultural relevance, and collaboration in teaching, research, service, and community engagement. It also offers real-world knowledge to prepare students to be successful professionals. During the school year, the college of education had a student body of approximately 4, 211 in which 3,301 females (78.39%) and 910 males (21.61%) were represented in the population. Of these students, 2,655 identified as White (63.05%), 900 as Black (21.37%), 459 as Hispanic (10.90%), and 197 as other (4.68%). Students enrolled in this college included 3, 286 undergraduates (78%) and 925 (22%) graduates in a number of different departments. Approximately 1,416 (34%) students were housed in the Elementary Education Department, 1,062 (25.5%) students in the Kinesiology & Health Science Department, 797 (19.1%) students in the School of Human Sciences, 558 (13.4%) students in the Human Services Department, and 334 (8%) in the Secondary Education & Educational Leadership Department. The demographic of the faculty within the college included 26 full professors (13.07%), 25 associate professors (12.56%), 14 teaching assistants (7.04%), 78 adjunct faculty (39.20%), 13 instructors (6.53%), 7 lecturers (3.52%), and 11 as other (5.53%). Of these faculty members, 110 were non-tenure track (55.28%), 55 were tenured (27.64%) and 34 were on-track (17.09%). About 106 of faculty had no terminal degree (53.27%) and 93 faculty had completed a terminal degree in their field of study (46.73%).

Description of Department. The selected department of this study has a faculty body of researcher-practitioners, who are invested in the preparation of special education teachers for elementary and secondary schools, preparation of persons for careers in rehabilitation, orientation and mobility, counseling and related human services, occupations serving persons with disabilities, speech language pathology and school psychology Faculty information was gathered using BANNER, JackFacts, and faculty activity reports (FARS),faculty vita, annual faculty activity reports, and course evaluation from archival data. During the academic year noted, the department consisted of 4 undergraduate majors (Communication Disorder, Deaf & Hard of Hearing, Pre-Audiology, and Rehabilitation Services), 2 undergraduate and graduate majors (Orientation & Mobility and Special Education), 5 graduate majors (Clinical Mental Health Counseling, Rehabilitation Counseling, School Counseling, Speech Language Pathology, and Students Affairs & Higher Education), and one graduate and doctoral major (School Psychology). The department had a student body of approximately 558 with 511 being female (91.58%) and 47 being male (8.42%). Of these students, 381 identified as White (68.28%), 89 as Black (15.95%), 59 as Hispanic (10.57%), and 29 as other (5.20%). Students enrolled in this department included 324 undergraduates and 234 graduates. The demographic of the faculty for the department included 6 full professors (12.77%), 3 teaching assistants (6.38%), 16 adjunct faculty (34.04%), 6 assistant professors (12.77%), 6 associate professors (12.77%), 3 instructors (6.38%), 1 lecturer (2.13%), and 6 as other (12.77%). Of these faculty members, 26 were non-tenure track

(55.32%), 11 were tenured (23.40%) and 10 were on-track (21.28%). About 26 of the faculty had no terminal degree (55.32%) and 21 faculty had completed a terminal degree in their field of study (44.68%).

Results

Pearson Product Correlation analyses were used to examine relationships between all pair of variables to test for correlations that met the criteria of multicollinearity. Multicollinearity occurs when there are strong and statistically significant relationships between pairs of independent variables to be included in a regression analysis, increases the standard error of the coefficients, which increases the probability of Type I error. This is the failure to accept the null hypothesis when it is actually true, which make some variables statistically insignificant when they should be significant. The criterion for determining multicollinearity includes a cutoff value of r = .80 (Thompson, Kim, Aloe, & Becker, 2017). The correlations that met the criterion for multicollinearity were the number of faculty publications and faculty work load. Faculty work load had a significant correlation with tenure, terminal degree, and faculty rank that met the criteria of multicollinearity. Publications had a significant correlation with presentations that met the criteria for multicollinearity. Tenure, terminal degree, faculty rank, and presentation were included because these variables were supported by the literature; therefore, faculty work load and publication were excluded from the regression. Backward multiple regression analyses were performed to examine the degree to which the independent variables (i.e., faculty demographic, academic degree, academic status, professional

engagement, grading patterns, and course variables) predict the three dependent

variables: faculty cumulative course evaluation, course evaluation standard deviation, and

response rates.

To reduce the chances of obtaining Type I error in the use of several multiple

regressions using the same independent variables, in which a false-positive result is

identified (rejecting the null hypothesis when you should not; likelihood of a significant

result by chance), a Bonferroni correction is used to address the potential problem

(Armstrong, 2014). The new p-value will be the alpha-value (α original = .05) divided by

the number of comparisons (3): (α altered = .05/3) = .016. To determine if any of the

models in the regression are statistically significant, the p-value must be $p < .016$. The

Bonferroni correction indicates statistical significance within this study given the number

of regressions performed with the same body of independent variables.

Dependent Variable: Overall Course Evaluation

The Pearson Product Correlations for the total sample, indicated a relationship

between overall course evaluations and the following variables: gender ($r = -.173$, $p =$

.004), ethnicity ($r = -.149$, $p = .012$), year of terminal degree ($r = .218$, $p < .001$),

institution ($r = .156$, $p = .009$), presentations ($r = .162$, $p = .006$), #Bs assigned ($r = -.162$,

$p = .022$), #Fs assigned ($r = -.267$, $p = .020$), course evaluation standard deviation ($r = -$

.478, $p < .001$), and response rate ($r = .246$, $p < .001$). Findings indicate that faculty with

higher course evaluations tend to be female, identify as racial majority, received terminal

degree recently, attended a research institution, engage in more professional

presentations, assign fewer grades of Bs and Fs, have a smaller standard deviation, and receive higher course evaluation response rate (see Table 1 for the Pearson Product correlation matrix indicating the relationships between all pair of variables). No other significant relationships were found between overall course evaluation and the independent variables.

      Research question one. To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree, status of institution where degree was awarded), faculty academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), faculty professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in overall course evaluation ratings? Results of the backward regression analysis indicated that the year of terminal degree significantly predicted the variance in overall course evaluations, $F(1, 27) = 3.063$, $p < .000$, $R^2 = .602$ (see Table 2). Sixty percent of the variance in overall course evaluations was explained by the faculty members' year of terminal degree completion.

Dependent Variable: Standard Deviation

      The Pearson Product Correlations for the total sample, indicated a relationship between standard deviation of course evaluations and the following variables: gender ($r = .217$, $p < .001$), ethnicity ($r = .196$, $p = .001$), year terminal degree ($r = -.189$, $p = .001$),

institution ($r = -.117$, $p = .050$), presentations ($r = -.224$, $p < .001$), #Bs assigned ($r = .220$, $p = .002$), #Fs assigned ($r = .260$, $p = .024$), course type ($r = -.194$, $p = .001$), course level ($r = -.194$, $p = .001$), and overall course evaluation ($r = -.478$, $p < .001$). Findings indicate that faculty with greater standard deviations in course evaluations tend to be male, identify as racial minority, completed their terminal degree earlier, attended a non-research institution, engage in fewer presentations, assign more grades of Bs and Fs, teach lecture courses, teach lower level courses, and have a lower overall course evaluation (see Table 1 for the Pearson Product correlation matrix indicating the relationships between all pair of variables). No other significant relationships were found between standard deviations and the independent variables.

Research question two. To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree, status of institution where degree was awarded), faculty academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), faculty professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in standard deviations of course evaluation? Results of the backward regression analysis indicated that faculty ethnicity and year of terminal degree significantly predicted the variance in standard deviations of course evaluations, $F(2, 26) = .462$, $p < .000$, $R^2 = .550$ (see Table 2). Fifty-five percent of the variance in course evaluation

56

standard deviations was explained by the faculty members' ethnicity and year of terminal degree completion.

Dependent Variable: Response Rate

The Pearson Product Correlations for the total sample, indicated a relationship between response rates and the following variables: gender ($r = -.181$, $p = .002$), faculty rank ($r = -.129$, $p = .030$), faculty work load ($r = -.118$, $p = .046$), publications ($r = -.160$, $p = .007$), # As assigned ($r = .223$, $p < .001$), # Cs assigned ($r = -.252$, $p = .005$), # Ds assigned ($r = -.288$, $p = .033$), # Fs assigned ($r = -.324$, $p = .005$), course delivery ($r = .176$, $p = .003$), course type ($r = -.247$, $p < .001$), and overall course evaluation ($r = .246$, $p = < .001$). Findings suggest that faculty with higher response rates tend to be female faculty, has a lower rank position, are part-time faculty, has fewer publications, assign more grades of As, assign fewer grades of Cs, Ds, Fs, teach more online courses, teach lecture courses, and receive higher course evaluations (see Table 1 for the Pearson Product correlation matrix indicating the relationships between all pair of variables). No other significant relationships were found between response rates and the independent variables.

Research question three. To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree, status of institution where degree was awarded), faculty academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), faculty professional

engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the within course variance in response rates? Results of the backward regression analysis indicated that the number of Bs, Ds, and Fs assigned predicted the variance in response rates, $F (3, 25) = .120$, $p = .018$, $R^2 = .245$ (see Table 2). However, this did not meet the Bonferroni adjusted $p \leq .016$, although it did approach significance.

Table 1

Pearson Product Correlation Matrix

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. Gender	-	.100	-.346**	.117*	.380**	.326**	.476**	.439**	.222**	.470**	-.301**	-.292**	.134	.123	.092	.271**	-.300**	.035	-.020	-.173**	.217**	-.181**
2. Ethnicity	.100	-	.153**	.099	.080	.157**	-.006	.112	.062	.096	-.189**	.080	.121	.105	.234	-.099	-.038	-.156**	.002	-.149*	.196**	-.028
3. Year Terminal Degree	-.346**	.153**	-	.004	.063	-.031	-.344**	-.042	-.260**	-.462**	.027	.082	-.145*	.092	.028	-.087	.244**	-.123*	-.134*	.218**	-.189**	.021
4. Institution	.117*	.099	.004	-	.521**	.740**	.567**	.659**	.436**	.376**	.154**	-.075	.034	.118	.148	.097	-.050	-.069	.158**	.156**	-.117*	-.114
5. Tenure	.380**	.080	.063	.521**	-	.733**	.771**	.857**	.433**	.406**	.172**	-.136*	-.054	.039	.126	.152	-.038	-.050	.085	.028	.013	.098
6. Terminal Degree	.326**	.157**	-.031	.740**	.733**	-	.790**	.877**	.495**	.479**	.202**	-.158**	.013	.044	.106	.164	.031	.010	.161**	.003	.055	-.112
7. Faculty Rank	.476**	-.006	-.344**	.567**	.771**	.790**	-	.872**	.585**	.610**	.167**	-.224**	-.016	-.051	-.018	.136	-.100	.010	.156**	.162**	.042	-.129*
8. Faculty Work Load	.439**	.112	-.042	.659**	.857**	.877**	.872**	-	.493**	.465**	.170**	-.192**	.021	.073	.126	.162	.002	-.034	.154**	-.045	.016	-.118*
9. Presentations	.222**	.062	-.260**	.436**	.433**	.495**	.585**	.493**	-	.814**	.068	-.187**	-.144*	-.057	.043	-.040	-.082	.075	.407**	.162**	-.224**	-.088
10. Publications	.470**	.096	-.462**	.376**	.406**	.479**	.610**	.465**	.814**	-	.010	-.153*	-.026	-.014	.092	.268**	-.235**	.151*	.283**	-.002	-.057	-.160**
11. Licenses	-.301**	-.189**	.027	.154**	.172**	.202**	.167**	.170**	.068	.010	-	.166**	-.219**	-.282**	-.102	-.152	.215**	-.021	.136*	.103	-.112	-.021
12. # of A's	-.292**	.080	.082	-.075	-.136*	-.158**	-.224**	-.192**	-.187**	-.153*	.166**	-	.258**	.027	.146	-.030	.020	-.221**	-.276**	.035	.220**	-.022
13. # of B's	.134	.121	-.145*	.034	-.054	.013	-.016	.021	-.144*	-.026	-.219**	.258**	-	.580**	.471**	.360**	-.291**	-.200**	-.393**	-.162*	.009	-.252**
14. # of C's	.123	.105	.092	.118	.039	.044	-.051	.073	-.057	-.014	-.282**	.027	.580**	-	.750**	.116	-.262**	-.121	-.183*	-.045	.104	-.288**
15. # of D's	.092	.234	.028	.148	.126	.106	-.018	.126	.043	.092	-.102	.146	.471**	.750**	-	.132	-.183	---	-.162	-.143	.104	-.324**
16. # of F's	.271**	-.099	-.087	.097	.152	.164	.136	.162	-.040	.268**	-.152	-.030	.360**	.116	.132	-	-.004	---	.132	-.267**	.260**	.176**
17. Course Delivery	-.300**	-.038	.244**	-.050	-.038	.031	-.100	.002	-.082	-.235**	.215**	.020	-.291**	-.262**	-.183	-.004	-	-.173**	.050	.107	.176**	.246**
18. Course Type	.035	-.156**	-.123*	-.069	-.050	.010	.010	-.034	.075	.151*	-.021	-.221**	-.200**	-.121	---	---	-.173**	-	.164**	-.055	-.194**	-.247**
19. Course Level	-.020	.002	-.134*	.158**	.085	.161**	.156**	.154**	.407**	.283**	.136*	-.276**	-.393**	-.183*	-.162	.132	.050	.164**	-	.002	-.194**	-.478**
20. Overall Course Evaluation	-.173**	-.149*	.218**	.156**	.028	.003	.162**	-.045	.162**	-.002	.103	.035	-.162*	-.045	-.143	-.267**	.107	-.055	.002	-	-.478**	.104
21. Standard Deviation	.217**	.196**	-.189**	-.117*	.013	.055	.042	.016	-.224**	-.057	-.112	.220**	.009	.104	.104	.260**	.176**	-.194**	-.194**	-.478**	-	.104
22. Response rate	-.181**	-.028	.021	-.114	.098	-.112	-.129*	-.118*	-.088	-.160**	-.021	-.022	-.252**	-.288**	-.324**	.176**	.246**	-.247**	-.478**	.104	.104	-
N	284	284	284	284	284	284	284	284	284	284	284	279	200	120	55	75	284	284	284	284	284	284

Note: Gender: coded: 1 = female, 2 = male; Ethnicity: coded: 1 = racial majority, 2 = racial minority;

Institution where degree was awarded: coded: 1 = non-research, 2 = research institution; Tenure: coded: 1 = non-tenure, 2 = tenure;

Terminal degree: coded: 1 = MA, 2 = PhD; Faculty rank: coded: 1 = adjunct, 2 = clinical/instructor, 3 = assistant, 4 = associate, 5 = full;

Faculty work load: coded: 1 = part-time 2 = full-time; Course delivery type: coded: 1 = none online (face-to-face), 2 = some online (hybrid), 3 = all online;

Course type: coded: 1 = lecture, 2 = practicum; Course level: coded: 1 = undergraduate (100-400 level), 2 = master (500 level), 3 = doctorate (600 level).

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 2
Hierarchical Regression Analysis for Overall Course Evaluation, Standard Deviation, and Response Rate

Predictor	Overall Course Evaluation					Standard Deviation					Response Rate				
	B	SE B	β	R¹	ΔR²	B	SE B	β	R¹	ΔR²	B	SE B	β	R¹	ΔR²
Initial Model				.851	.406				.871	.480				.770	.123
Gender	-.310	.278	-.373			.243	.147	.517			.215	.159	.549		
Ethnicity	.007	.315	.006			.254	.166	.439			-.153	.180	-.318		
Year Terminal Degree	.030	.014	.697			-.016	.007	-.641			.010	.008	.488		
Institution	.198	.234	.233			-.016	.124	-.033			-.041	.134	-.103		
Tenure	.149	.854	.177			-.485	.452	-1.021			-.736	.489	-1.859		
Terminal Degree	-.259	.528	-.304			-.034	.279	-.071			.365	.302	.907		
Faculty Rank	.122	.221	.381			.096	.117	.531			.116	.126	.769		
Presentations	-.010	.020	-.196			.012	.010	.424			.004	.011	.180		
Licenses	-.007	.036	-.038			.018	.019	.190			.004	.020	.054		
As	.015	.016	.292			-.014	.008	-.494			-.007	.009	-.306		
Bs	-.017	.022	-.307			.019	.012	.595			.026	.013	.987		
Cs	9.809	.036	.001			.020	.019	.366			-.009	.021	-.190		
Ds	.005	.066	.026			-.042	.035	-.384			-.017	.038	-.188		
Fs	.003	.072	.010			.011	.038	.055			-.090	.041	-.568		
Course Delivery	-.006	.108	-.013			.057	.057	.238			.076	.062	.379		
Final Model				.785	.602				.763	.550				.571	.245
Ethnicity						.178	.077	.308							
Year Terminal Degree	.034	.005	.785			-.019	.003	-.795							
Bs											.011	.006	.410		
Ds											-.046	.018	-.512		
Fs											-.073	.029	-.464		

A backward regression was conducted to examine variables associated with the variance of course evaluations. Analyses were performed to address the following three research questions:

1) To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree?, status of institution where degree was awarded), faculty academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), faculty professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in overall course evaluation ratings?

2) To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree, status of institution where degree was awarded), faculty academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), faculty professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in the standard deviation of the course evaluation?

3) To what degree do faculty demographics (i.e., race, gender), academic background (i.e., years since completion of terminal degree, status of institution where degree was awarded), faculty academic status (i.e., terminal degree, tenure vs. non-tenured, faculty rank, faculty work/class load), faculty professional engagement as noted on curriculum vitae (i.e., # presentations, # publications, # licenses/certifications), course grading patterns (i.e., #As, Bs, Cs, Ds, and Fs), and course characteristics (i.e., course level, instructional delivery type, and course type) predict the variance in response rates?

Overall Course Evaluation Analysis

The backward regression indicated that, the year of terminal degree significantly predicted the variance in overall course evaluations, $F (1, 27) = 3.063$, $p < .000$, $R^2 = .602$ and accounted for 60.2% of the variance. Faculty that received their degree recently obtained higher course evaluations. It is important to note that the regression successfully met Bonferroni correction p.

Standard Deviation Analysis

The backward regression indicated that, the ethnicity and year of terminal degree significantly predicted the variance in standard deviations of course evaluations, $F_{(2, 26)}$ = .462, $p < .000$, $R^2 = .550$ and accounted for 55% of the variance. Faculty that received their degree recently obtained smaller standard deviations, meaning less within-classroom variance in students' evaluations. The regression successfully met Bonferroni correction p.

Response Rate Analysis

The backward regression indicated that the number of Bs, Ds, and Fs assigned predicted the variance in response rates, $F_{(3, 25)}$ = .120, $p = .018$, $R^2 = .245$ and accounted for 24.5% of the variance in response rates. Faculty that assigned fewer D's and Fs's were likely to receive a higher response rate. It is important to note that the response rate failed to meet, but did approach Bonferroni correction p with .016, which warrants attention in future research. It is important to note that this model would have been significant with $p = .05$.

Chapter V

Discussion

The current study was designed to examine three research questions to aid in understanding the meaning of data associated with course evaluations as a measurement to provide information concerning teaching effectiveness, and a tool commonly used by administration during the decision of faculty members' promotion and tenure process. In the literature addressing student course evaluations, there is currently no analysis that examines all the variables simultaneously. This study provides a more in-depth examination of the meaning of course evaluations by examining: To what degree do faculty demographic, academic degree, academic status, professional engagement, grading patterns, and course variables predict the variance in university faculty cumulative course evaluation, course evaluation standard deviation, and response rate?

Information provided by results of the first research question, in which faculty academic background (the year of terminal degree) predict the variance in university faculty cumulative course evaluation, found that faculty that received their degree recently accounted for a significant amount of the variance in overall course evaluation. The results of the second research question, in which faculty academic background (the year of terminal degree) predict the variance in course evaluations standard deviations, found that faculty that received their degree recently accounted for a significant amount of variance in standard deviation. For the final research question, in which faculty

63

grading patterns (the number of As, Bs, Cs, Ds, and Fs assigned during the academic year course evaluations were collected) predict the variance in response rate, found that the number of Ds and Fs assign predict the variance in response rates, but did not reach a level of statistical significance.

This study concluded that faculty academic background was found to provide a significant contribution to the amount of variance in university faculty cumulative course evaluation and standard deviation within the backward regression model. This study identified no statistical significance between faculty demographic, academic status, professional engagement, grading patterns, and course variables. Previous research have combined these various factors and found it to have some effects to a degree on course evaluations that is important to consider when evaluating teaching effectiveness and in the tenure and promotion process (Bavishi et al., 2010; Johnson et al., 2013; McPherson, Jewell, & Kim, 2009; Isley & Singh, 2007; Mohan, 2011; Prince et al., 2007; Brockx et al., 2011; Young & Duncan, 2014). However, in the current study each of these variables are independent from one another to evaluate each predictor's relation to the criterion variable.

Prior research has shown that faculty academic background influences students' rating on course evaluations. Based on traditional belief that knowledge and experience are based on level of degree, it is implied that faculty year of terminal degree should result in higher ratings on course evaluations. Researchers have found that tenured instructors received significantly lower rating on their course evaluations compared to

non-tenured faculty (Johnson et al., 2013). These findings indicate that course evaluations are dependent upon faculty members' year of terminal degree. Faculty members that received their degree recently are commonly identified as non-tenured or adjunct faculty. These faculty members may possess characteristics such as productivity, interaction, and enthusiasm as they enter into the profession, which could influence higher evaluations (Stonebraker & Stone, 2015; Tang & Chamberlain, 2003). Non-tenure track faculty are hired to 'teach' and this is the task on which their evaluation is based and job security maintained. However, faculty, who have tenure-track appointments, must maintain optimal levels of teaching, scholarship, and service to maintain their position and positive annual evaluations. Although, the effects of this predictor on course evaluations have been rarely examined the findings in this current research indicate significant influence (Landrum, 2009). There are few studies that indicate faculty members' years in terminal degree, nonetheless this variable does influence student perceptions when rating faculty.

At this time little is known about the varied ratings of course evaluations within the same course, which makes these findings a critical contribution to the literature. It is important to note that understanding faculty variables associated with course evaluation standard deviations might be considered in the review and interpretation of students' cumulative ratings. These findings indicated that faculty that received their degree most recently tended to have smaller standard deviations. Results indicate smaller variance in students' ratings, which mean the students tended to have more similar perceptions of the

faculty members' teaching effectiveness. Although not, statistically significant, findings suggest that for racial-ethnic minority faculty obtained greater standard deviations, meaning more within-classroom variance in students' evaluations. Plausible reasons for these results could be contributed to students' relatedness to faculty. This study contributes to the growing body of research by focusing specifically on the importance of examining course evaluations standard deviations when evaluating teaching effectiveness rather than focusing primarily on the mean rating.

Examining the grading patterns and its influence on response rate has shown to contribute to the likelihood of students responding to the course evaluation at the end of each term. Though this variable did not reach significance in this model, it important to note the variance accounted for by grading patterns in relation to response rates. Faculty, who assign more Ds and Fs are likely to obtain lower response rates, which decrease the validity of the overall rating. These findings suggest that students, who received the grades D and F, are less likely to participate in the course evaluations than students, who receive higher grades (Hatfield & Coyle, 2013). In addition, it is likely that students, who receive lower grades than anticipated participate in course evaluation by giving the faculty lower ratings on the course evaluations due to their dissatisfaction with their received grade (Sinclair & Kunda, 2000). This may lead to bias ratings and inaccurate reflection of teaching effectiveness. Research has supported that completing evaluations during class time has also help increase response rates (Guder & Malliaris, 2013). This may increase the validity of response rates and the results. Further research is warranted

to consider similar findings, and it is recommended that academicians consider the influence of grading patterns when examining course evaluations for validity.

Limitations

Data in this correlational study were gathered from an academic unit housed in a college of education, during one academic school year at a university. The study was restricted to courses found in a human services-related academic unit, which also includes two courses designated as Core Curriculum. Student demographics were unavailable to analyze because the university course evaluation is given in a manner to protect students' anonymity. The literature noted that course evaluation patterns may differ based on location of the data collection (Chen & Hoshower, 2003). Studies at other institutions may differ in findings due to variance in university demographic (Chen &Yeager, 2011). Necessity of local research should be conducted periodically to understand cultural norms based on students and faculty practices, and to better understand the unique variables associated with what students consider teaching excellence.

This is also a theoretical analysis; therefore, multiple limitations may impact the interpretation of this study and the overall results. A theoretical analysis is often used in a research study to gather concepts to help understand, challenge, and form predictions based on prior findings in the literature (Jasso, 1988). In this study, the literature review only provides existing knowledge of course evaluation to guide and strengthen a theory about faculty characteristics, grading practices, and course types that is significantly

associated with students' rating of faculty cumulative course evaluation, course

evaluation standard deviation, and response rate. The findings of this theory also

highlight the key variables in course evaluations that should be further examine in an

exploratory analysis.

A primary limitation of this study is the large number of potential independent

variables in the backward regression. Psychologists need a theory behind the analysis, in

which a theory can be develop based on multiple variables. The problem is that there is

not enough support or existing knowledge to guide selection among the variables.

Therefore, a backward regression is responsible in the process of making decisions about

the significance of the selection. However, arguments tend to arise when using this

approach because of the value place on the computer judgment of the given sample rather

than the experimenter's logic of the analysis.

Another limitation of this study is the application of the Bonferroni corrections

when a large number of independent variables are being used. Having too many

Bonferroni corrections in a research design make it almost impossible to find anything

significant and creates more problems (Perneger, 1998). The likelihood of chance

increases when the independent variable contribution to the R^2 is determined by the

significance test. The problem is the interpretation of the number of tests performed and

the validity of the results when the tests for each independent variable are being

performed concurrently; there is a competition for significance. There is also an increase

chance of type II error, which is the probability of accepting the null hypothesis when it

is false but failing to reject it. It is likely that some variables will appear insignificant when it is deemed important to the findings. It is important to note that even with the Bonferroni correction two models were found to be significant; more construct variables would have reached significance.

Limitations of the research design, backward regression, also include problems with the generalizability given that the computer is making decisions about the current sample. It is likely that the independent variables in this sample may not be applicable to other samples in a population. The difference between the variables may change between each variable across populations. Also, when the independent variables are correlated with each other, the problem is often linked to a variable being removed when labeled insignificant; this usually occurs because of insufficient contribution in the model. It is important to note that this study cannot be generalized to other institutions and may have limited generalizability to other university demographics.

Implications

Despite numerous myths and misconceptions regarding academicians and students' perceptions of course evaluation, these findings suggest implications for student affairs, administration, and faculty members' reviews from a higher education policy perspective. Course evaluations information gives insight into the students' perception of teaching effectiveness in relation to faculty instructional performance and background, and aid in the process of promotion and secured employment (Milanowski, 2017). Findings has shown that the faculty year of terminal degree can be accounted for 60.2%

of variance in cumulative course evaluation and 55% of variance in course evaluation standard deviation. These results are important for those in positions of authority to consider in review of university faculty course evaluations, especially when used in decision of promotion and tenure and merit raises. While this raised minimum concern for faculty that receive their degree recently in the study, faculty that received their degree earlier may have a greater risk in receiving lower course evaluations (Wollert & West, 2000). For example, faculty that received their degree recently as they enter into the field may be perceived as more enthusiastic and interactive based on students' perceptions therefore receiving higher ratings. In the case of waning overall, course evaluations, senior faculty, in preparation for post-tenure review, might be reminded of the need to recover the enthusiasm and engagement with students.

Other variables such as faculty demographics (the ethnicity and sex of faculty members) and faculty academic background (the year of their terminal degree and if the degree was awarded from a research vs. non-research institution) are those over which the instructor has limited to no control (Smith, 2017). When reviewing the course evaluation standard deviation, one might consider the distribution of response in which the class rated the course evaluation. For example, a course evaluation that has a small deviation suggest that the students within the class rated the same. While a course evaluation that has a large standard deviation, indicates a wide variance in students' response and that faculty should consider reviewing the reason for the difference across ratings. This information is important to interpret because the degree to which having a

greater or smaller standard deviation depends on the median of students' ratings and faculty cumulative course evaluation. For faculty that received positive response (higher overall course evaluation), a smaller deviation is desired, while faculty that may have obtained negative response (lower overall course evaluation) would favor a greater standard deviation.

Examining grading patterns could also inform interpretations of faculty course evaluations. Findings from this current study have shown the number of Ds and Fs assign has accounted for 24.5% of the variance in response rates. It can be hypothesized that low response rates might be associated with students who were dissatisfied with their grade and chose not to respond (Simpson & Siguaw, 2000; Williams & Ceci, 1997). These results are important to interpret because it suggests that the grades faculty assign can influence higher or lower response rates from students which could also affect their cumulative course evaluation. Previous research has found that faculty who assign higher grades can expect to have higher course evaluations than faculty who assign lower grades to students (Brockx et. al, 2011; Krautmann & Sander, 1999; McPherson, 2006). Grade inflation is one of many concerns of reviewers of course evaluations due to pressure on faculty to give higher grades to obtain a better evaluation and improve chances at promotion and tenure. It is important that faculty examine the likelihood of such influence on their course evaluations, especially if these grading patterns are essential in considering during the promotion and tenure process. To decrease the probability of this error, faculty might provide access to the course evaluations before final grades are

posted, or give students the opportunity to complete course evaluations in class to improve response rates. Administrators, who evaluate dossiers for promotion and tenure, might focus more so on examining the relationship between each faculty cumulative course evaluation and the standard deviation received. Furthermore, policies might include a required mid semester and final semester course evaluation, and an attainable percentage of response rates to ensure a valid score of course evaluations before it is included in the promotion and tenure process.

These findings can assist faculty members to understand and explain their course evaluation, more specifically their instructional performance, students' perceptions, and the quality of the course content (King & Fraser, 2016). This may help faculty in addressing improvement in course content and student engagement. It is also recommended that faculty take advantage in early course evaluations to adjust the course and teaching strategies that best enrich the students' learning environment (McGowan & Osguthorpe, 2011). A general training about course evaluations for students, faculty, and administrators could reduce the misinterpretation of what the instrument is measuring and more accurately interpret course evaluations (Franklin, 2001; Lewis, 2001). As a result, an increase in validity and reliability on course evaluations may occur and successfully train effective evaluators of faculty instruction, improve the quality of learning in courses, enrich self-improvement of teaching performance, and a justified promotion and tenure process.

Future Research and Recommendations

Due to an absence of comparable studies, replication is essential for this analysis. Future research should be conducted to explore the relationship between course evaluations and these variables all at one time across colleges within universities. Also, periodic local research to understand criteria students use in course evaluation and response rates is warranted to determine if this is consistent or if other findings may be indicated (Abrami et al., 2007; Ginns et al., 2007). Findings suggest recommendations for administration and student affairs practices to develop programming in orientations that prepare students as trained evaluators. Administrators should use multiple means to assess course instruction given the potential for gender bias, especially in the cases of annual review, promotion, tenure, and merit (Baldwin & Blattner, 2003; Subbaye & Vithal, 2017; Zabaleta, 2007).

The variance within a classroom evaluation has never been examined, but may warrant special attention given that students exposed to the same experience were found to see things differently due to their very own performance in the class and the degree to which the faculty member varies grading. Future research might examine the median of the students' response to identify the areas of concern and to distinguish between characteristics of students who were most satisfied and those who were not. Variables associated with classroom characteristics (i.e., teaching strategies, lecture style vs. discussion style) might be included in the model. This kind of information would provide faculty with well-defined points to address for improvement. Faculty can use this

feedback information to adjust instructional performance, course content, and manner of engagement to mediate negative responses that might be attributed to experiences that have nothing to do with course instruction. Although ethnicity did not reach significance in this model, future research is warranted to examine the relationship between this variable and the standard deviation as it did approach significance. Therefore, faculty should review the standard deviation when reviewing dossier for promotion and tenure.

Additionally, in regards to response rates, it is recommended that faculty provide prompts and reminders to students to complete course evaluations at the end of each semester (Anderson et al., 2006). It is suggested that faculty explain the importance of course evaluations in improving the course design, state that anonymity of responses, send reminders to encourage participation, and possibly propose incentives to promote completion of course evaluation prior to grades being posted. Faculty may also ensure that students know their grades throughout the semester so they might be aware of the final grade long before the end, receive advisement from the faculty about how to improve, and have time to be clear about the reason for their failure and that the reason resides with them, given that other students have been successful with the same assignments. To the leaders in the field of School Psychology, it is imperative to contribute in the investigation of course evaluations given that this is a critical component of professional development evaluation within an educational setting (Williams, 2010). Although this is a theoretical study, an exploratory analysis of ethnicity and year of terminal degree in future studies may contribute to identifying the significance of these

specific variables in students' course evaluation. Collection of students' characteristics and the influence on course evaluations is especially important given the variance in standard deviations for racial-ethnic minorities which suggests that students exposed to the same instructor see the experiences differently, probably due to individual variables having to do with attitudes and experiences not included in most studies of course evaluations. Future research might also include mediation analysis, path analyses and/or structural equation modeling, student demographics in the model, and data collection would span across academic units.

Conclusion

Overall, this study contributes to the field as it questioned the meaning of course evaluations and its influence on teaching effectiveness and promotion and tenure decisions. Though replication is needed, these findings may assist faculty in interpreting and understanding course evaluation so that an accurate reflection of both teaching efficacy and effectiveness is detected. Furthermore, a reliable course evaluation will render a fairer assessment of evaluation of promotion and tenure for those who are devoted and exceed expectations in within academic programs in higher education.

References

Abrami, P. C., D'Apollonia, S., & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. *In R. P. Perry & J. C. Smart (Eds.), The scholarship of teaching and learning in higher education: An evidence-based perspective (pp. 385–445). Dordrecht, The Netherlands: Springer.*

Ahmadi, M., Helms, M. M., & Raiszadeh, F. (2001). Business students' perceptions of faculty evaluations. *International Journal of Educational Management*, *15*(1), 12-22. doi:10.1108/09513540110366097

Alhija, F. N. A., & Fresko, B. (2009). Student evaluation of instruction: What can be learned from students' written comments? *Studies in Educational Evaluation*, *35*(1), 37-44.

Al Kuwaiti, A., AlQuraan, M., & Subbarayalu, A. V. (2016). Understanding the effect of response rate and class size interaction on students' evaluation of teaching in a higher education. *Cogent Education*, *3*(1), 1204082. doi:10.1080/2331186X.2016.1204082

Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., . . . Bray, M. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching, 52*(4), 134-141. doi:10.3200/CTCH.52.4.134-141

Allen, M. (1996). Research productivity and positive teaching evaluations: Examining the relationship using meta-analysis. *Journal of the Association for Communication, 2*, 77–96.

Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States*. Babson Survey Research Group and Quahog Research Group. Retrieved from http://www.onlinelearningsurvey.com/reports/changingcourse.pdf

Anderson, J., Brown, G., & Spaeth, S. (2006). Online student evaluations and response rates reconsidered. *Innovate: Journal of Online Education*, *2*(6), 5.

Armstrong, J. S. (1998). Are student ratings of instruction useful? *American Psychologist, 53*(11), 1223-1224. doi:10.1037/0003-066X.53.11.1223

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502-508. doi:10.1111/opo.12131

Baldwin, T., & Blattner, N. (2003). Guarding against potential bias in student evaluations: What every faculty member needs to know. *College Teaching*, *51*(1), 27-32. doi:10.1080/87567550309596407

Ballantyne, R., Borthwick, J., & Packer, J. (2000). Beyond student evaluation of teaching: Identifying and addressing academic staff development needs. Assessment & Evaluation in Higher Education, 25(3), 221-236. doi:10.1080/713611430

Baruch, Y. (1999). Response rate in academic studies-A comparative analysis. *Human Relations*, *52*(4), 421-438. doi:10.1177/001872679905200401

Basow, S. A. (2000). Best and worst faculty members: Gender patterns in students' choices. *Sex roles*, *43*(5), 407-417. doi:10.1023/A: 1026655528055

Bavishi, A., Madera, J. M., & Hebl, M. R. (2010). The effect of faculty member ethnicity and gender on student evaluations: Judged before met. *Journal of Diversity in Higher Education*, *3*(4), 245.

Benton, S. L., & Cashin, W. E. (2012). Student ratings of teaching: A summary of research and literature. Manhattan, KS: IDEA. Retrieved from https://www.cte.cornell.edu/documents/blog/idea-paper_50.pdf

Beran, T. N., & Rokosh, J. L. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science*, *37*(2), 171-184. doi:10.1007/s11251-007-9045-2

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education*, *30*(6), 593-601. doi:10.1080/02602930500260688

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, *17*(1), 48-62.

Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.

Boysen, G. A. (2016). Using student evaluations to improve teaching: Evidence-based recommendations. *Scholarship of Teaching and Learning in Psychology*, *2*(4), 273-284. doi:10.1037/stl0000069

Brockx, B., Spooren, P., & Mortelmans, D. (2011). Taking the grading leniency story to the edge. The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability*, *23*(4), 289-306. doi:10.1007/s11092-011-9126-2

Cain, J., Romanelli, F., & Smith, K. M. (2012). Academic entitlement in pharmacy education. *American Journal of Pharmaceutical Education*, *76*(10), 189.

Centra, J. A. (1983). Research productivity and teaching effectiveness. *Research in Higher Education*, *18*(4), 379-389.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, *44*(5), 495-518. doi:10.1023/A:1025492407752

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

Chan, C. K., Luk, L. Y., & Zeng, M. (2014). Teachers' perceptions of student evaluations of teaching. *Educational Research and Evaluation*, *20*(4), 275-289. doi:10.1080/13803611.2014.932698

Chau, P. (2010). Online higher education commodity. *Journal of Computing in Higher Education*, *22*(3), 177-191.

Chen, C. M., & Chen, M. C. (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. *Computers & Education*, *52*(1), 256-273. doi:10.1016/j.compedu.2008.08.005

Chen, Q., and J. L. Yeager. 2011. Comparative study of faculty evaluation of teaching practice between Chinese and U.S. institutions of higher education. *Frontiers of Education in China,* 6 (2), 200–226

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & evaluation in higher education*, *28*(1), 71-88. doi:10.1080/02602930301683

Cho, D., & Cho, J. (2017). Does more accurate knowledge of course grade impact teaching evaluation?. *Education Finance and Policy*, *12*(2), 224-240

Clayson, D. E. (1999). Students' evaluation of teaching effectiveness: Some implications of stability. *Journal of Marketing Education*, *21*(1), 68-75. doi:10.1177/0273475399211009

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, *31*(1), 16-30. doi:10.1177/0273475308324086

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*, *21*(2), 101-112. doi:10.2753/MER1052-8008210201

Crumbley, L., Henry, B. K., & Kratchman, S. H. (2001). Students' perceptions of the evaluation of college teaching. *Quality Assurance in Education*, *9*(4), 197-207. doi:10.1108/EUM0000000006158

Darwin, S. (2012). Moving beyond face value: re-envisioning higher education evaluation as a generator of professional knowledge. *Assessment & Evaluation in Higher Education*, *37*(6), 733-745. doi:10.1080/02602938.2011.565114

Denson, N., Loveday, T., & Dalton, H. (2010). Student evaluation of courses: what predicts satisfaction? *Higher Education Research & Development*, *29*(4), 339-356. doi:10.1080/07294360903394466

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, *29*(5), 611-623. doi:10.1080/02602930410001689171

Dunrong, B., & Fan, Meng (2009). On student evaluation of teaching and improvement of the teaching quality assurance system at higher education institutions. *Chinese Education & Society*, *42*(2), 100-115. doi:10.2753/CED1061-1932420212

Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*, *11*(1), 37-46. doi:10.1108/09684880310462074

Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, *43*(4), 483-501. doi:10.1023/A:1015579817194

Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education*, *26*(3), 227-298.

Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education*, *28*(4), 291-329.

Filak, V. F., & Sheldon, K. M. (2003). Student psychological need satisfaction and college teacher-course evaluations. *Educational Psychology*, *23*(3), 235-247. doi:10.1080/0144341032000060084

Flores, K. L., Matkin, G. S., Burbach, M. E., Quinn, C. E., & Harding, H. (2012). Deficient critical thinking skills among college graduates: Implications for leadership. *Educational Philosophy and Theory*, *44*(2), 212-230. doi:10.1111/j.1469-5812.2010.00672.x

Fosnacht, K., Sarraf, S., Howe, E., & Peck, L. K. (2017). How important are high response rates for college surveys? *The Review of Higher Education*, *40*(2), 245-265. doi:10.1353/rhe.2017.0003

Franklin, J. (2001). Interpreting the numbers: Using a narrative to help others read student evaluations of your teaching accurately. *New Directions for Teaching and Learning*, *2001*(87), 85-100. doi:10.1002/tl.10001

Gaillard, F. D., Mitchell, S. P., & Kavota, V. (2011). Students, faculty, and administrators perception of students evaluations of faculty in higher education business schools. *Journal of College Teaching & Learning (TLC)*, *3*(8).

Gray, P. J. (1992). *A national study of research universities on the balance between research and undergraduate teaching*. Center for Instructional Development, Syracuse University, Syracuse, NY.

Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education*, *32*(5), 603-615. doi:10.1080/03075070701573773

Giovannelli, M. (2003). Relationship between reflective disposition toward teaching and effective teaching. *The Journal of Educational Research*, *96*(5), 293-309. doi:10.1080/00220670309597642

Goodman, J., Anson, R., & Belcheir, M. (2015). The effect of incentives and other instructor-driven strategies to increase online student evaluation response rates. *Assessment & Evaluation in Higher Education*, *40*(7), 958-970. doi:10.1080/02602938.2014.960364

Gray, M., & Bergmann, B. R. (2003). Student teaching evaluations. *Academe*, *89*(5), 44-46.

Guder, F., & Malliaris, M. (2013). Online course evaluations response rates. *American Journal of Business Education (Online)*, *6*(3), 333.

Haladyna, T. M., & Amrein-Beardsley, A. (2009). Validation of a research-based student survey of instruction in a college of education. *Educational Assessment, Evaluation and Accountability*, *21*(3), 255-276. doi:10.1007/s11092-008-9065-8

Hammonds, F., Mariano, G. J., Ammons, G., & Chambers, S. (2017). Student evaluations of teaching: improving teaching quality in higher education. *Perspectives: Policy and Practice in Higher Education*, *21*(1), 26-33. doi:10.1080/13603108.2016.1227388

Hatfield, C. L., & Coyle, E. A. (2013). Factors that influence student completion of course and faculty evaluations. *American Journal of Pharmaceutical Education*, *77*(2), 27.

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research*, *66*(4), 507-542. doi:10.2307/1170652

Heckert, T. M., Latier, A., Ringwald-Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: is it possible to" buy" better evaluations through lenient grading? *College Student Journal*, *40*(3), 588.

Heckert, T. M., Latier, A., Ringwald, A., & Silvey, B. (2006). Relation of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness. *College Student Journal*, *40*(1), 195.

Ho, A. K., Thomsen, L., & Sidanius, J. (2009). Perceived academic competence and overall job evaluations: Students' evaluations of African American and European American professors. *Journal of Applied Social Psychology*, *39*(2), 389-406. doi:10.1111/j.1559-1816.2008.00443.x

Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations: What all faculty should know. *College Teaching*, *49*(1), 26-31. doi:10.1080/87567550109595842

Hodges, L. C., & Stanton, K. (2007). Translating comments on student evaluations into the language of learning. *Innovative Higher Education, 31*(5), 279-286. doi:10.1007/s10755-006-9027-3

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

Isely, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *The Journal of Economic Education*, *36*(1), 29-42. doi:10.3200/JECE.36.1.29-42

Jasso, G. (1988). Principles of theoretical analysis. *Sociological Theory, 6*(1), 1-20. doi:10.2307/201910

Jaquett, C. M., VanMaaren, V. G., & Williams, R. L. (2016). The effect of extra-credit incentives on student submission of end-of-course evaluations. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 49-61. doi:10.1037/stl0000052

Johnson, M. D., Narayanan, A., & Sawaya, W. J. (2013). Effects of course and instructor characteristics on student evaluation of teaching across a college of engineering. *Journal of Engineering Education, 102*(2), 289-318. doi:10.1002/jee.20013

Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology, 3*, 44. doi:10.3389/fpsyg.2012.00044

King, L., & Fraser, D. (2016). Who is learning what from student evaluations of teaching? *Waikato Journal of Education*, *11*(2),. doi:10.15663/wje.v11i2.312

Kinney, D. P., & Smith, S. P. (1992). Age and teaching performance. *The Journal of Higher Education*, *63*(3), 282-302.

Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. W. (2010). Student evaluations of teaching: perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, *15*(6), 623-636. doi:10.1080/13562517.2010.491911

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review*, *18*(1), 59-63.

Landrum, R. E. (2009). Are there instructional differences between full-time and part-time faculty? *College Teaching*, *57*(1), 23-26. doi:10.3200/CTCH.57.1.23-26

Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review*, *27*(4), 417-428. doi:10.1016/j.econedurev.2006.12.003

Lewis, K. G. (2001). Making sense of student written comments. *New Directions for Teaching and Learning*, *2001*(87), 25-32. doi:10.1002/tl.25

Lewis, K. G. (2001). *Techniques and strategies for interpreting student evaluations*. Jossey-Bass.

Liu, O. L. (2012). Student evaluation of instruction: In the new paradigm of distance education. *Research in Higher Education, 53*(4), 471-486. doi:10.1007/s11162-011-9236-1

Love, D. A., & Kotchen, M. J. (2010). Grades, course evaluations, and academic incentives. *Eastern Economic Journal*, *36*(2), 151. doi:10.1057/eej.2009.6

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291-303. doi:10.1007/s10755-014-9313-4

Malouff, J. M., Reid, J., Wilkes, J., & Emmerton, A. J. (2015). Using the results of teaching evaluations to improve teaching: A case study of a new systematic process. *College Teaching*, *63*(1), 3-7. doi:10.1080/87567555.2014.956681

Marlin Jr, J. W. (1987). Student perception of end-of-course evaluations. *The Journal of Higher Education*, *58*(6), 704-716.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal*, *17*(2), 219-237.

Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complementary, antagonistic, or independent constructs? *The Journal of Higher Education*, *73*(5), 603-641. doi:10.1353/jhe.2002.0047

Marsh, H. W., & Overall, J. U. (1981). The relative influence of course level, course type, and instructor on students' evaluations of college teaching. *American Educational Research Journal*, *18*(1), 103-112. doi:10.2307/1162533

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, *52*(11), 1187-1197. doi:10.1037/0003-066X.52.11.1187

Martz, E. (2013). Enough is enough! Handling multicollinearity in regression analysis. Retrieved from http://blog.minitab.com/blog/understanding-statistics/handling-multicollinearity-in-regression-analysis

McGowan, W. R., & Osguthorpe, R. T. (2011). Student and faculty perceptions of effects of midcourse evaluation. *To Improve the Academy*, *29*(1), 160-172.

McPherson, M. A. (2006). Determinants of how students evaluate teachers. *The Journal of Economic Education*, *37*(1), 3-20.

McPherson, M. A., & Jewell, R. T. (2007). Leveling the playing field: Should student evaluation scores be adjusted? *Social Science Quarterly*, *88*(3), 868-881. doi:10.1111/j.1540-6237.2007.00487.x

McPherson, M. A., Jewell, R. T., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal, 35*(1), 37-51. doi:10.1057/palgrave.eej.9050042

Mentzer, G. A., Cryan, J., & Teclehaimanot, B. (2007). Two peas in a pod? A comparison of face-to-face and web-based classrooms. *Journal of Technology and Teacher Education*, *15*(2), 233.

Milanowski, A. (2017). Lower performance evaluation practice ratings for teachers of disadvantaged students: Bias or reflection of reality? *AERA Open*, *3*(1),. doi:10.1177/2332858416685550

Mohan, N. (2011). On the use of non-tenure track faculty and the potential effect on classroom content and student evaluation of teaching. *Journal of Financial Education*, 29-42.

Morrison, K. (2013). Online and paper evaluations of courses: A literature review and case study. *Educational Research and Evaluation, 19*(7), 585-604. doi:10.1080/13803611.2013.834608

Nargundkar, S., & Shrikhande, M. (2014). Norming of student evaluations of instruction: Impact of non-instructional factors. *Decision Sciences Journal of Innovative Education*, *12*(1), 55-72. doi:10.1111/dsji.12023

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, *27*(2), 187-198. doi:10.1080/02602930220128751

National Association of School Psychologists Practice Model. (2017). Who are school psychologists. Retrieved from https://www.nasponline.org/about-school-psychology/who-are-school-psychologists

Nowell, C., Gale, L. R., & Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in Higher Education*, *35*(4), 463-475. doi:10.1080/02602930902862875

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301-314. doi:10.1080/02602930701293231

Onwuegbuzie, A. J., Witcher, A. E., Collins, K. M., Filer, J. D., Wiedmaier, C. D., & Moore, C. W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis. *American Educational Research Journal*, *44*(1), 113-160. doi:10.3102/0002831206298169

Parker, J. 2008. Comparing research and teaching in university promotion criteria. *Higher Education Quarterly,* 62 (3), 237–251. doi:10.1111/j.1468-2273.2008.00393.x

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ : British Medical Journal*, *316*(7139), 1236–1238.

Prince, M. J., Felder, R. M., & Brent, R. (2007). Does faculty research improve undergraduate teaching? An analysis of existing and potential synergies. *Journal of Engineering Education*, *96*(4), 283-294. doi:10.1002/j.2168-9830.2007.tb00939.x

Regoniel, Patrick A. (November 11, 2012). Example of a research using multiple regression analysis. *Simply Educate Me*. Retrieved from http://simplyeducate.me/2012/11/11/example-of-a-research-using-multiple-regression-analysis/

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

Richardson, J.T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education, 30*(4), 387-415. doi:10.1080/02602930500099193

Riffell, S., & Sibley, D. (2005). Using web-based instruction to improve large undergraduate biology courses: An evaluation of a hybrid course format. *Computers & Education*, *44*(3), 217-235. doi:10.1016/j.compedu.2004.01.005

Roche, L. A., & Marsh, H. W. (2002). Teaching self-concepts in higher education: reflecting on multiple dimensions of teaching effectiveness. *Teacher Thinking, Beliefs and Knowledge in Higher Education*.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368-384. doi:10.1016/j.compedu.2007.05.016

Sandoval, J., & Love, J. A. (1977). School psychology in higher education: The college psychologist. *Professional Psychology, 8*(3), 328-339. doi:10.1037/0735-7028.8.3.328

Schiekirka, S., & Raupach, T. (2015). A systematic review of factors influencing student ratings in undergraduate medical education course evaluations. *BMC Medical Education*, *15*(1), 30. doi:10.1186/s12909-015-0311-8

Sheehan, E. P., & DuPrey, T. (1999). Student evaluations of university teaching. *Journal of Instructional Psychology*, *26*(3), 188.

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, *22*(3), 199-213. doi:10.1177/0273475300223004

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, *26*(11), 1329-1342.

Smith, B. P. (2007). Student ratings of teaching effectiveness: An analysis of end-of-course faculty evaluations. *College Student Journal*, *41*(4), 788.

Smith, K., & Welicker-Pollak, M. (2008). What can they say about my teaching? Teacher educators' attitudes to standardised student evaluation of teaching. *European Journal of Teacher Education*, *31*(2), 203-214. doi:10.1080/02619760802000248

Sonner, B. S. (2000). A is for adjunct: Examining grade inflation in higher education. *Journal of Education for Business*, *76*(1), 5-8. doi:10.1080/08832320009599042

Spencer, K. J., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, *27*(5), 397-409. doi:10.1080/0260293022000009285

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*(4), 598-642. doi:10.3102/0034654313496870

Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, *53*(11-12), 779-793. doi:10.1007/s11199-005-8292-4

Sproule, R. (2000). Student Evaluation of Teaching: Methodological Critique. *Education Policy Analysis Archives*, *8*, 50. doi:10.14507/epaa.v8n50.2000

Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*, *9*, 2014. doi:10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Stark-Wroblewski, K., Ahlering, R. F., & Brill, F. M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: Supplementing student evaluations of teaching with pre–post learning measures. *Assessment & Evaluation in Higher Education*, *32*(4), 403-415. doi:10.1080/02602930600898536

Stehle, S., Spinath, B., & Kadmon, M. (2012). Measuring teaching effectiveness: Correspondence between students' evaluations of teaching and different measures of student learning. *Research in Higher Education*, *53*(8), 888-904. doi:10.1007/s11162-012-9260-9

Stonebraker, R. J., & Stone, G. S. (2015). Too old to teach? The effect of age on college and university faculty members. *Research in Higher Education*, *56*(8), 793-812. doi:10.1007/s11162-015-9374-y

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. doi:10.1177/1745691616650284

Stuber, J. M., Watson, A., Carle, A., & Staggs, K. (2009). Gender expectations and on-line evaluations of teaching: Evidence from RateMyProfessors.com. *Teaching in Higher Education*, *14*(4), 387-399. doi:10.1080/13562510903050137

Stupans, I., McGuren, T., & Babey, A. M. (2016). Student evaluation of teaching: A study exploring student rating instrument free-form text comments. *Innovative Higher Education, 41*(1), 33-42. doi:10.1007/s10755-015-9328-5

Subbaye, R., & Vithal, R. (2017). Teaching criteria that matter in university academic promotions. *Assessment & Evaluation in Higher Education*, *42*(1), 37-60. doi:10.1080/02602938.2015.1082533

Summers, J. J., Waigandt, A., & Whittaker, T. A. (2005). A comparison of student achievement and satisfaction in an online versus a traditional face-to-face statistics class. *Innovative Higher Education*, *29*(3), 233-250. doi:10.1007/s10755-005-1938-x

Tang, T. L. P., & Chamberlain, M. (2003). Effects of rank, tenure, length of service, and institution on faculty attitudes toward research and teaching: The case of regional state universities. *Journal of Education for Business*, *79*(2), 103-110. doi:10.1080/08832320309599097

Thomas, D. A. (2014). Searching for significance in unstructured data: Text mining with leximancer. *European Educational Research Journal*, *13*(2), 235-256. doi:10.2304/eerj.2014.13.2.235

Thomas, E. H., & Galambos, N. (2004). What satisfies students? Mining student-opinion data with regression and decision tree analysis. *Research in Higher Education*, *45*(3), 251-269. doi:10.1023/B:RIHE.0000019589.79439.6e

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*, *39*(2), 81-90. doi:10.1080/01973533.2016.1277529

Titus, J. J. (2008). Student ratings in a consumerist academy: Leveraging pedagogical control and authority. *Sociological Perspectives*, *51*(2), 397-422. doi:10.1525/sop.2008.51.2.397

Tucker, B. (2014). Student evaluation surveys: Anonymous comments that offend or are unprofessional. *Higher Education, 68*(3), 347-358. doi:10.1007/s10734-014-9716-2

Tuckman, B. W. (2002). Evaluating ADAPT: A hybrid instructional model combining web-based and classroom components. *Computers & Education*, *39*(3), 261-269. doi:10.1016/S0360-1315(02)00045-3

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.

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, *23*(2), 191-211. doi:10.1080/0260293980230207

Wilkinson, L. (1979). Tests of significance in stepwise regression. *Psychological Bulletin*, *86*(1), 168-174. doi:10.1037/0033-2909.86.1.168

Williams, B. B. (2010). National Association of School Psychologists Principles for Professional Ethics. *School Psychology Review*, *39*(2), 302-319.

Williams, W. M., & Ceci, S. J. (1997). How am I doing? Problems with student ratings of instructors and courses. *Change: The Magazine of Higher Learning*, *29*(5), 12-23. doi:10.1080/0009138970960233

Willits, F., & Brennan, M. (2017). Another look at college student's ratings of course quality: data from Penn State student surveys in three settings. *Assessment & Evaluation in Higher Education*, *42*(3), 443-462. doi:10.1080/02602938.2015.1120858

Winchester, M. K., & Winchester, T. M. (2012). If you build it will they come? Exploring the student perspective of weekly student evaluations of teaching. *Assessment & Evaluation in Higher Education*, *37*(6), 671-682. doi:10.1080/02602938.2011.563278

Wollert, M. H., & West, R. F. (2000). Differences in student ratings of instructional effectiveness based on the demographic and academic characteristics of instructors. Bowling Green, KY: Mid-South Educational Research Association.

Wongsurawat, W. (2011). What's a comment worth? How to better understand student evaluations of teaching. *Quality Assurance in Education, 19*(1), 67-83. doi:10.1108/09684881111107762

Wright, S. L., & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education*, *37*(6), 683-699. doi:10.1080/02602938.2011.56327

Young, S., & Bruce, M. A. (2011). Classroom community and student engagement in online courses. *Journal of Online Learning and Teaching*, *7*(2), 219.

Young, S., & Duncan, H. E. (2014). Online and face-to-face teaching: How do student ratings differ? *Journal of Online Learning and Teaching*, *10*(1), 70.

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, *12*(1), 55-76. doi:10.1080/13562510601102131

Appendix

College of Education Course Evaluation

Course:
Instructor:

Your input is very important in the assessment of both course content and instructor effectiveness. Although your instructor will be able to view the names of students who participate in the survey, the ratings you are about to complete will be anonymous and will not be available to your instructor until after final grades have been posted.

Course/instructor evaluation data will be used for both course and program improvement and for instruction evaluation purposes. Such information has a significant impact on tenure, promotion, pay, and retention decisions. Therefore, please be thoughtful, thorough, and accurate in completing the evaluation. The information from the survey also aids the individual faculty member in course planning and improvement of teaching.

Please make your selections from the following.

	Very Good	Good	Average	Poor	Very Poor	No Opinion
1. The course as a whole was	○	○	○	○	○	○
2. The clarity of course objectives was	○	○	○	○	○	○
3. The usefulness and relevance of course content to course objectives was	○	○	○	○	○	○
4. The organization of the course was	○	○	○	○	○	○
5. The instructor's overall effectiveness in teaching was	○	○	○	○	○	○
6. The variety and appropriateness of learning activities in this course was	○	○	○	○	○	○
7. The instructor's ability to present and explain course content was	○	○	○	○	○	○
8. The instructor's use of examples and illustrations was	○	○	○	○	○	○
9. The instructor's enhancement of students' interest in course content was	○	○	○	○	○	○
10. My confidence in the instructor's knowledge was	○	○	○	○	○	○
11. The instructor's enthusiasm was	○	○	○	○	○	○
12. My interest level in each session (either online or face-to-face) was	○	○	○	○	○	○
13. The availability of extra help when requested was	○	○	○	○	○	○
14. The use of course time (either online or face-to-face) was	○	○	○	○	○	○
15. The instructor's interest in student learning and success was	○	○	○	○	○	○
16. The amount I learned during the course was	○	○	○	○	○	○
17. The evaluation and grading techniques (tests, papers, projects, etc.) were	○	○	○	○	○	○
18. The amount and appropriateness of assigned work was	○	○	○	○	○	○
19. The clarity of my responsibilities and requirements was	○	○	○	○	○	○
20. Additional Comments:						

VITA

Tyesha Stewart graduated from Stephen F. Austin State University with a
Bachelor of Arts degree in Psychology and a minor in Child and Family Development in
December of 2014. In spring of 2015, Tyesha enrolled at Stephen F. Austin State
University to earn a Master of Arts degree in School Psychology. The program of School
Psychology has provided Tyesha with the opportunity to gain valuable experience with
psychoeducational assessments, academic and behavioral interventions, behavior
analysis, counseling, ethics and laws, multiculturalism, systems-level services, special
education eligibility, and assessing children representing a wide range of ages,
populations, and needs within several school districts. She also had the opportunity to
work on multiple projects, including identifying bullying assessment tools, factors
associated with school dropout rates, improving student's behavior using self-monitoring,
effectiveness of classroom management, and personality and perception of online
courses. In the future, Tyesha plans to start a private practice along with an early
childhood center, and work toward faculty appointment in higher education settings.

Permanent Address: 1931 Blue Sage Dr.
                                     Lancaster, TX 75146

Publication Manual of the American Psychological Association, Sixth Edition

This thesis was typed by Tyesha D. Stewart