2018

Student evaluations of teaching: perceptions of faculty knowledge and their relation to learning

Donald L. Gruendler Jr.

Pepperdine University

Graduate School of Education and Psychology

STUDENT EVALUATIONS OF TEACHING: PERCEPTIONS OF FACULTY KNOWLEDGE AND THEIR RELATION TO LEARNING

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Education in Organizational Leadership

by

Donald L. Gruendler, Jr.

October, 2018

Doug Leigh, Ph.D. – Dissertation Chairperson

This dissertation, written by

Donald L. Gruendler, Jr.

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Doctoral Committee:

Doug Leigh, Ph.D., Chairperson

Andrew Harvey, Ed.D.

Elio Spinello, Ed.D.

TABLE OF CONTENTS

LIST OF TABLES

DEDICATION

This manuscript is dedicated to my parents, Donald and Marilla Gruendler. Although you were on this earth for a short time, I am forever grateful for the sacrifices you made in an effort to provide me with the best opportunities possible. Not only did you stress my needing to obtain an education; but also, you both encouraged me to concurrently follow my passion for music, which altogether led to my studies at Pepperdine and this resulting research. I love you.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Doug Leigh, my committee chair. You are much more than an advisor or instructor; you are a friend and I am very fortunate to have had your encouragement and friendship throughout this program. I am also so very grateful to my two other committee members, Dr. Andrew Harvey, and Dr. Elio Spinello, each of whom worked so well together. While a committee of three may have initially seemed minute, your perspectives, edits, support, and friendship has been invaluable. Not only during this dissertation process; but also within my coursework in such classes as Quantitative Research, Consultancy Project, and E-Learning and Design. As a result of our collaborations and your collective guidance, I am now a better researcher, educator, author, and person.

In addition to those already mentioned above, I must also give thanks to the many others who have enriched my experience at Pepperdine including, but not limited to: Dr. Jack McManus for an amazingly rewarding Comprehensive Exam experience; Dr. Vance Cesar for his inspiration, mentorship, and friendship; Dr. Farzin Majidi for my "million-dollar degree"; Dr. June Schmieder-Ramirez for allowing me to follow my muse within each portion of the EDOL Program; LTD: Patty Coaley, Carole Bennett and Sundra Ryce for constantly challenging me to be better and deliver more; and my entire Fall 2015 cohort for enriching my life and coursework.

Most importantly, I must honor my incredible wife Hope for her love, support, patience, and friendship; my sons DeeGee and Graham for their immense joy, inspiration, and sense of purpose; My "Gram" June Tiffin for all of her support; and everyone who has helped, inspired, or aided in this research study.

VITA

EDUCATION

Master of Music 1996-1998
- Wayne State University, Detroit, MI. The Graduate-Professional Full Scholarship (Magna Cum Laude)

Bachelor of Music 1993-1996
- Berklee College of Music, Boston, MA. (Cum Laude)

EXPERIENCE

2018-Present

Vice President of Education, Guitar Center Inc.
- Assisting the organization with its growth, operations, P&L, academic vision, pedagogical direction, program development, curricular and media materials, interdepartmental communication.
- $45M+ fiscal management, oversight of 300 locations, 1500 instructors, and 30,000+ students.

2015-2018

President, MI College of Contemporary Music
- $60M+ fiscal management, oversight of two campuses, multiple residence halls and 1400+ students.
- Milestones included: Initiated WASC candidacy, expanded to a new Nashville, TN campus, increased Hollywood campus enrollment by 10%, increased revenue by 22%, reduced cost per student start by 19%, increased international student population to 40%, founded MI Online (the college's first online offerings) and facilitated the acquisition of MI's first residence halls in Hollywood, CA.

PRIVATE

President, DLG Studio Services, Inc. 2002-Present
- Oversight of all aspects of music licensing for film, TV, games and web. Clients include: NBC, Universal, Disney, HBO, United Artists, Warner Brothers, DreamWorks, CBS, FX, Sony, NPR and EA. Projects include: Pirates of the Caribbean, Star Trek, Transformers, Money Wise, ReMax on the Blvd, and Chicago Overcoat.

Published Author, Self 2004-Present
- Authorship of 25+ periodicals, ten textbooks, five instructional DVDs and three iOS applications for a wide range of domestic and international publishers. Leadership, musical instruction and business related topics are covered in detail.

SPEAKING

Guest Lecturer UNESCO Creative Cities Network (UCCN), Berlin, Germany 2015
- A personal guest lecture invitation initiated by UNESCO to promote cooperation among cities which recognize creativity as a major factor in their urban development. Topics included instructional design, creativity in all facets of education and how these items impact a given community.

Guest Lecturer: Japanese External Trade Organization (JETRO), Hamamatsu, Japan 2014
- A personal guest lecture invitation initiated by the Japanese External Trade Organization to foster economic and educational partnerships between Japan, the city of Hamamatsu, and Hollywood, California. Also worked alongside the Mayor of Hamamatsu in numerous panel discussions.

ABSTRACT

The majority of student evaluations of teaching (SET) related studies repeatedly consider matters related to the creation and validity of an assessment tool, as well as the validity, and reliability of SET scores. Not only to determine the usefulness of teaching; but also, the possible sources of student biases related thereto as well (Hofman & Kremer, 1980; Abrami & Mizener, 1983; Tollefson et al., 1989). However, limited research studies have considered SET and their relation to student learning outcomes. Therefore, the purpose of this study was to identify what relationships, if any, exists between the grade undergraduate college students' predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. Also, this study also examined what relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

The population for this study consisted of 344 undergraduate college students enrolled during the spring 2018 quarter at a small private college of music, located in Hollywood, California. Students' predicted grade was collected via a document that contained a detailed outline of the course grade percentage standards alongside a single question survey: "What Grade do you expect to earn in this course?" Students' perceptions of faculty knowledge was measured via 15 questions, covering an instructor's subject matter knowledge (e.g., "My instructor understands the topics at a high-level ") and inquiries from the knowledge of students' understanding header (e.g., "My instructor is familiar with my prior knowledge in this subject area"). Measuring for students' affect toward faculty was accomplished through the administering of an eight-item survey assessing respondents' affect toward the instructor, which

included items such as the value of instruction and the like. Extant data regarding students actual earned grade, and overall SET evaluation was collected on the last day of classes, the college's Office of the Registrar.

A detailed investigation of the Wald statistic of the individual relationships revealed that none of the various grades students' predicted to earn in a given had a significant impact on the prediction of the different students' actual earned grade in a given course after controlling the effect of students' perception of faculty knowledge and affect toward faculty. However, the undergraduate college students' actual grade earned was significantly related to students' perceptions of faculty knowledge ($F(4, 339) = 2.86$, $p = 0.02$), and their affect toward instructor ($F(4, 339) = 77.27$, $p < 0.001$). The findings mentioned above are further reinforced by the Post-Hoc test findings too. Specifically, that undergraduate students' who earn a higher-level grade concurrently rate their faculty member as having high knowledge, versus those students who earn a low final course grade. Also, undergraduate college students who earn higher-level grades also have a higher rating of affect toward instructor than those undergraduate college students who concluded their studies with a lower grade earned.

Chapter I: Introduction

Overview

Students at most U.S. colleges and universities partake in an end of term evaluation in which they rate their faculty on a set of fixed variables that are intended to measure teaching effectiveness alongside the quality of course (Kolitch & Dean, 1999). This type of evaluation is labeled student evaluations of teaching (SET), and the results thereof are employed to make critical judgments regarding faculty course assignments, tenure, and possible positional promotion alongside the systemic demonstration of institutional effectiveness to various higher education accrediting agencies.

However, for at least the last few decades, teacher and course evaluations have commonly been little more than a bureaucratic exercise, often failing to help administrators, teachers and students recognize either excellence or mediocrity within teaching methods and student learning (Stark & Freishtat, 2014). As such (and in this context) evaluation has been comprised of missed opportunities for giving teachers valuable feedback that could help them improve both their practice and the student achievement of academic course objectives and outcomes.

While academic evaluation systems may not be perfect, they do come from honorable intentions and seek to ensure that said classes and faculty members offer a first-class scholastic experience to students. However, when this same evaluation system merely administers questionnaires to students without adequately linking them to student learning outcomes or efficiently appraising the data, the resulting information can often be skewed or misinterpreted. Also, and since evaluations are not usually a mandatory occurrence in the classroom, it is also

difficult to adequately measure how well professors (or their courses) rate in the eyes of their students as a whole. Possibly most concerning is that learners do not always assess faculty solely on their instructional effectiveness. Thus, many researchers have discovered that students frequently utilize bias factors such as physical appeal, personality, and professional accomplishment alongside their course based achievement to formulate their teacher and course evaluations (Witt, Jerome, and Burdalski, n.d.).

With the items mentioned above in mind, and although student evaluations of teaching (SET) have been comprehensively examined, the veracity of their current assessment processes has not been verified. One issue has been the correlation between SET and student learning outcomes. Thus, through this research project, I sought to examine work product related to student evaluations of teaching in collegiate undergraduate studies and more specifically, if said evaluations were expressly associated with student learning. It is also my hope that other institutions may utilize this resulting information as an entrance into the subject matter, thereby having the potential to help creative institutions improve their quality of evaluation, instruction, and course offerings.

SET Background

Not many subjects within the field of undergraduate collegiate instruction have been as well studied, documented, and long argued as SETs. Over 90 years ago, Remmers & Brandenburg (1927) published the first article on student's attitudes toward instructors and their teaching. Three decades later, the University of Washington became one of the earliest colleges to conduct an official evaluation and analysis of teaching and student objective outcomes as well (Guthrie, 1954). An extensive amount of research occurred during the 1970s, and much of the

resulting findings have become well known and implemented. By the mid-1990s, 80% of college campuses used some form of SET within their instructional evaluation processes (Seldin, 1999). In many Academic Affairs Offices, student appraisals of teaching are often the most significant and often, the single gauge of instructional capability (Wilson, 1998). Seldin (1999) recounted an academic administrator stating, "If I trust one source of data on teaching performance, I trust the students" (p. 15).

As the SET device is easy to fill out and takes little faculty or class time, it is also quite popular amongst academic administrators, and its consequential scores are the most common method employed to assess instruction (Cashin, 1999; Clayson, 2009; Seldin, 1999; Wilson, 1998). The calculated median for a faculty member's SET ratings may also appear equitable due to its numerical rating scale. In addition, comparing an individual teacher's score to all-encompassing program (i.e., departmental) averages is simple. Regrettably, many of the surveys used in faculty evaluation processes are custom made by an administrator from a given institution, which calls their structure, validity, and reliability into question. It is also unclear whether or not institutions should utilize SETs as the lone source of information regarding instructional quality, and the effectiveness of SET questions and methods of interpretation persist (Pounder, 2007). Faculty also believe that student evaluations are an integral component of administrative decision making (Barth, 2008; Beran & Rokosh, 2009b). Likewise, colleges have also not agreed on a unified definition of the skills and characteristics needed to demonstrate teaching excellence (Arreola, 2007, p. 98).

Challenges of SET. Although most academics seem to agree that there is a need for instructional evaluation, the problem is that they do not concur on whether or not the present-day

methods and instruments utilized have a valid application for this purpose (Marsh, 2007). Both the SET research and their related assessment tools are hampered by the same challenge – i.e., no consensus exists on (or commonly accepted definition of) what "good" teaching is, nor has a completely agreeable standard of teaching effectiveness been established (J. V. Adams, 1997). Further, and since disparate courses also do not follow a standard method of evaluating student performance, dispersing assignments, administering exams or determining a grade, quantifying learning can be difficult. As many undergraduate courses are also quite interpretive, grades are not perceived as one's full measure of learning and can be a poor gauge of student progress (Beleche, Fairris, & Marks, 2012).

Non-mandatory SET response rates are also commonly low and fluctuate between 30% and 50% (Al-Maamari, 2015; Arnold, 2009). Dommeyer, Baum, Hanna, and Chapman (2004) further indicated that the average response rates of synchronous in-class instructor assessments reach nearly 70%, while asynchronous digital SET deliver 29% participation. Further, the impact of low response rates on SET scores is less well-known. That is, waning participation within the SET process continues to create faculty skepticism and distrust, thereby enabling critics to further question the legitimacy of SET (Macfadyen, 2016).

Other items also call the validity of SETs into question as well. For example, teachers perceived as passionate, friendly, cooperative, impartial, knowledgeable, and successful were rated as more effectual instructors than those who did not have these qualities attributed to them (Barth, 2008; Hills, Naegle, & Bartkus, 2009; Stark & Freishtat, 2014). The "leniency hypothesis" or otherwise-titled "paradox of rigor" assumes that instructors who employ a lenient grading scheme may gather more positive SETs and as a result -- achieve superior overall

evaluation scores too (Gump, 2007; Heckert, Latier, Ringwald-Burton, & Drazen, 2006). Many studies also agree that students are perfectly situated to give feedback regarding their experience in a course, which – in a music school, such as the setting of the present study – involves factors that guide teaching effectiveness including a faculty member's linguistic clarity, writing legibility, musical performance level, professional notoriety and possibly the instructor's availability during office hours and the like (Stark & Freishtat, 2014). They can also mention whether they feel more inspired about the subject matter (upon conclusion of the course) and if said class motivated them to take a related or follow-up serial course. However, faculty and administrators question whether or not students have the scholastic background, academic achievement or life experience to properly gauge teaching performance or a faculty member's knowledge within said subject matter (Marsh, 1987). Finally, and notwithstanding the ability of the instructor, the attributes of a course may have an effect on SETs too, which are normally comprised of a class (large or small) and classroom atmosphere, i.e. modern or improved spaces or rooms with a good deal of noise, less-than-perfect lighting and uncomfortable seating (Hill & Epps, 2010).

Furthermore, investigative projects have included (but are not necessarily limited to) studies regarding: the consistency of diverse student populations rating a given faculty member on the same SET form (e.g., Abrami, d'Apollonia, & Cohen, 1991; Braskamp and Ory, 1994; Centra, 2003; Ory, 2001); the uniformity of students SET responses who have enrolled in two (or more) concurrent or subsequent courses with the same instructor (e.g., Braskamp and Ory, 1994; Centra, 1993; Marsh, 2007; Marsh and Dunkin, 1992; Overall and Marsh, 1980); and

if SETs measure a student's biases toward and perceptions of a given faculty member and the resulting course environment (Boring, Ottoboni, & Stark, 2016).

Statement of the Problem

Through the systematic deployment and administration of student ratings via teaching questionnaires, higher education institutions have devoted precious resources to the processes of faculty and course evaluation. However, the resulting data collected and analyzed is a narrow means to evaluate one's breadth, depth, overarching quality and value of instruction. While there are advantageous uses for the resulting above-mentioned statistical SET figures, Ory and Ryan (2001) share that some institutional practitioners inadvertently misuse said information for purposes not originally intended. The process of administering student evaluations of teaching has been argued to have become meaningless and only deployed because of being mandated by a given institution (Ory & Ryan, 2001). Further, the current body of SET research has been largely confined to reliability and validity studies, rather than focusing how these items may be related to student learning outcomes or employed to improve instruction overall (Penny, 2003).

Purpose of Study

The purpose of this study was to identify what relationships, if any, exists between the grade undergraduate college students' predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. In addition, this study also examined what relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

This quantitative, quasi-experimental study examined students' estimates of their final course grade, their actual earned grade, and their sentiments related to faculty expertise and the degree of student learning in both the affective and cognitive domains. Data regarding these items were collected cross-sectionally at two points in time. Data regarding perceptions of faculty knowledge (a control variable) and affect toward faculty (a control variable) are at the level of interval measurement while predicted grade (a predictor variable) and actual earned grade (the outcome variable) in the course are categorical variables.

Research Questions. This study employed the following research questions:

1. What relationship, if any, exists between the grade undergraduate college students' predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty?
2. What relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET?

Hypotheses. It is essential to comprehend the effect of classroom instruction through students' perceptions, which influence student outcomes more than via the witnessed quality of teaching patterns and behaviors (Waxman & Huang, 1997). Higher student rankings on evaluation instruments may be granted to those faculty with more perceived expertise within a given area of study (Nowell, Gale, & Handley, 2010). In addition, students who anticipate earning a high grade have been found to be more likely to produce high grades and deliver superior faculty evaluation scores (Barth, 2008; Beran & Rokosh, 2009b; Beran & Violato, 2006; Beran & Violato, 2009; Nowell et al., 2010; Serdyukova, Tatum, & Serdyukov, 2010). With

these items in mind, and although the literature mentioned above supports a positive relationship between SET and student learning – i.e., college students' perceptions of faculty expertise, their predicted grade, and actual earned grade, I developed the following hypotheses for this study:

Null Hypothesis 1: The relationship between college students' predicted grade and actual earned grade is either negatively attenuated by students' perceptions of faculty expertise and their affect toward faculty, or not attenuated at all.

Alternative Hypothesis 1: It is hypothesized that the relationship between college students' predicted grade and actual earned grade is positively attenuated by students' perceptions of faculty expertise and their affect toward faculty.

Null Hypothesis 2: No relationship exists between college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

Alternative Hypothesis 2: A relationship exists between college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

Operational definitions. This study employed the following operational definitions:

- Predicted Grade — A student-generated prediction of the grade they believe they will earn in a given course. This student forecast was collected via a document that contains a detailed outline of the course grade percentage standards (see Table 3) alongside a single question survey: "What Grade do you expect to earn in this course?"
- Earned Grade — A faculty generated grade that a given student earned in a particular course. This data regarding a student's actual earned grade was collected on the last day of classes via MI's Office of the Registrar. May also be referred to as actual earned grade.

- Faculty Knowledge — This term refers instructors' proficiency at constructing circumstances favorable to student comprehension involving both tacit and explicit knowledge together, which is considered a distinctive form of knowledge (Feldman, 1986; Nonaka & Takeuchi 1996). Student perceptions of faculty knowledge was measured using a limited version of the Students' Perceptions of Faculty Knowledge (SPFK) instrument (Shih & Chuang, 2013).
- Affect Toward Faculty — The student respondents' affect toward a given instructor (i.e., positive feeling, attitude, beliefs toward and overall appraisals of said instructor). Measuring for students' affect toward faculty was accomplished through the Affective Learning Scale-Abbreviated (Mansson, 2014).
- SET Evaluation — The overarching terminology employed in this study to denote the dispersal, collection, evaluation, and analyzation of the SPFK instrument and Affective Learning Scale – Abbreviated.

Key definitions. This study employed the following key definitions:

- Administrator — A college program chair, dean, or vice-president whose responsibilities include managing and evaluating faculty.
- College or University — An accredited educational institution that awards undergraduate and graduate degrees.
- Faculty — A person who is credentialed to teach by the institution and is currently employed to do so at the collegiate level.
- Process — Multiple linear actions that lead to a final product; process and practice are key definitions utilized to disperse, collect and SET forms to assess instructional quality.

- Student Evaluations of Teaching (SET) — A widespread phrase that outlines the utilization of questionnaires or rating forms that students complete either synchronously or asynchronously to assess instructors. This phrase is similar to the following, which are also used in instructional evaluation literature:
    - Teacher Rating Forms (TRFs),
    - Student Evaluations of Faculty (SEF)
- Evaluation — The process of observing a matter and rating it, based on its significant features (Kiefer, 1994). The purpose of evaluation, regardless of subject matter, is to ascertain the present value of the subject according to the defined criteria to improve its quality in the future (Hajdin & Pažur, 2012).
- Learning — The active process in which learners construct new ideas or concepts based upon their current/past knowledge. The learner selects and transforms information, constructs hypotheses, and makes decisions, relying on a cognitive structure to do so. Cognitive structure (i.e., schema, mental models) provides meaning and organization to experiences and allows the individual to stretch past the information provided (Bruner, 2009).
- Successful Teacher — An individual that is proficient at constructing circumstances favorable to learning. Cohen (1981) affirmed, “Most researchers in this area agree that student learning is the most important criterion of teaching effectiveness” (p. 283).
- Value— To score, rate or scale in effectiveness, significance, or worth. (Merriam-Webster’s Collegiate Dictionary, 2017)
- Quality — To consider or rate highly (Merriam-Webster’s Collegiate Dictionary, 2017)

Setting. The population for this study consisted of students enrolled at Musicians Institute (MI), College of Contemporary Music, located in Hollywood, California. This institution is based on a quarterly academic system and allows for incoming quarterly student (i.e., new) enrollment alongside year-round, full-time study. Founded in 1977, the College is accredited by the National Association of Schools of Music (NASM), and provides Non-Credit, Professional Certificate, Associate of Arts, Bachelor's and Master's offerings. The National Association of Schools of Music (NASM) is recognized as one of the leading music professional associations in the world. Founded in 1924, the National Association of Schools of Music (2016) seeks to establish a uniform understanding among higher education music programs through the development of processes dedicated to the granting course credit, as well as for maintaining and improving threshold standards for the granting of degrees and other credentials. Further, this association of 659 institutions also seeks to facilitate and support student learning, development and effective musical practice (National Association of Schools of Music, 2016).

NASM states that music is best learned through practice and experience (not necessarily credential alone). In keeping with this, MI's College of Music has internationally recognized professional performing and recording faculty onsite and they have also been nationally recognized by the Council of Arts Accrediting Associations (CAAA) both for employing best practices in the area of student learning and development, as well as for demonstrating successful assessment and evaluation practices, which were foci of this study.

Importance of study. Although student evaluations of teaching (SET) have been comprehensively examined, the validity of their results has not been verified. One such highly

debated issue has been the correlation between SET and learning, which does alter the conduct of both faculty and students who deem that a relationship between grades and the evaluation process exist (Clayson, 2009). There is confirmation that this belief transforms faculty and student comportment as well (Marsh, 1987; Moore & Trahan, 1998; Redding, 1998; Ryan, Anderson, & Birchler, 1980; Simpson & Siguaw, 2000). As a result, I sought to demonstrate a positive relationship between the degree of student perception of faculty expertise and the degree of learning among undergraduate college students.

With the above-mentioned items in mind, this study adds to the body of literature related to SET. Specifically, the relationship between course, student, students' perceptions of faculty, and student ratings of teaching. While the findings within this study may not be unanimously applicable to the collegiate population at large, said results can function as additional data to be contemplated in measuring the impact of students' perceptions of faculty knowledge and their relation to learning and grade outcomes. In addition, this research employs numerous variables that have not been widely used in earlier studies (e.g. letter grades earned in music institutions, students predicted grade, and students' perception of faculty knowledge).

Assumptions. The researcher made three assumptions in this study, and they are as follows:

1. The institutional participants would be willing to partake in the study and would provide accurate and documentation related to the assessment initiatives for which they held on their campus.
2. All student participants would be truthful, and accurate in his or her participation throughout the research.

3. Student participant memories and recollections would be accurate, even over the span of time that had occurred since the evaluation initiative took place (i.e., the beginning of a given course).

Limitations. Formerly, the researcher was the Vice President of Academic Affairs at MI College of Music (Musicians Institute), and until January 15th, 2018 also served as MI's President. As such, said researcher was partially responsible for planning and implementing the student evaluation of teaching initiatives alongside the resulting annual reviews of faculty. The author of said study was also proportionately involved in institutional and academic assessment initiatives, such as establishing student learning outcomes and working to institute college-wide student learning outcomes as a part of a project team for the institution's WASC Candidacy and resulting accreditation process. Further, the researcher is also a notable educator in the field of contemporary music education with twelve books authored and another six co-authored. With these items in mind, those academic personnel, faculty and student participants at Musicians Institute, MI College of Music may have been reluctant to share information they perceived to be negative or that they believed may have adversely affected the institution, division, department, future accreditation, or themselves.

Further, select courses such as those that required public student musical performances or a good amount of student in-class participation may have caused some individuals to experience higher levels of anxiety than those courses that were lecture and homework based alone. As a result, it may have been difficult to determine the impact these feelings may have had on this study's student evaluations of teaching, including students' perceptions of faculty knowledge and their affect toward faculty. These anxieties may have also had an associated impact on a given

individual's weekly assignment and final course grades as well. As a result of said course types, the findings of this research study cannot be applied uniformly to the entire undergraduate populace throughout the world.

Delimitations. It was beyond the scope of this study to look at all types of colleges and universities. As such, the researcher focused on student evaluations of teaching at Musicians Institute, a small private undergraduate degree-granting music college, located in Hollywood, California. The population was limited to first through fourth-year undergraduate students who have completed a maximum of 135 total quarterly units and were enrolled in a minimum of 6 quarterly units during the Spring 2018 quarter on Musicians Institute's Hollywood campus. Thus, the scope of this research did not include General Education courses offered offsite via the college's various articulation agreements with regionally accredited institutions, for which second, third or fourth-year baccalaureate students may be enrolled.

As data collection occurred during MI's spring quarter, the timeframe of conducting said research was also another key restriction imposed by the researcher. That is, this study excluded SET from the college's summer, fall, and winter quarters.

Although much of the current SET literature has focused on student biases towards faculty; faculty biases of the SET administrative review process; impacts of classroom environments; and attributes of a given course; this research study sought to solely evaluate student learning outcomes, while controlling for SET items such as students' perceptions of faculty knowledge and their affect toward faculty. Learner, class, and teacher characteristics have been considered in past studies and as a result, were also not investigated in this research.

Further, and given the venue of this research study, both lecture and musical performance courses were included in this study. Per both Fiske (1977) and Kaiser (1998), no relationship has been found between an evaluators reliability in adjudicating student performances and their performing ability as measured by applied musical grades. However, each grade type was treated equally, and the researcher did not account for such items. It was also outside the reach of this study to analyze the use of student evaluations by collegiate administrative staff such as Deans and Program Chairs as well.

Summary

This chapter provided an introduction to the research study, which began with an overview of student evaluations of teaching (SET) in higher education. Not only did said discussion lead into the following challenges affecting SET involving (but not limited to) measuring teaching effectiveness and student learning outcomes; but also, the overarching foundation for the study was presented in detail as well. These items included a statement of the problem, the purpose of the study, research questions, research design, rationale for exercise and the importance of the study. The chapter concluded with a list of assumptions and key limitations of said study. The purpose of this chapter was to provide an introduction to the subject matter and a rationale for the research study.

With the items mentioned above in mind, Chapter II features a review of literature, including (but not necessarily limited to), the purpose of SET, challenges of measuring teaching effectiveness, impact to academic freedom, faculty and student characteristics, administration of SET, attributes of course, response rates, instrumentation and new methods of SET utilization. Chapter III examines the research design employed in this study, while Chapter IV presents the

results and alongside the discussion thereof. Finally, the Chapter V concludes this scholarship, while concurrently expanding on the items as mentioned earlier to include suggestions for future research and the like.

Chapter II: Review of Literature

Overview

Students at most American colleges complete some form of an end of semester evaluation in which they rate their instructor on a set of fixed variables that are intended to capture and measure effectual teaching alongside course quality (Kolitch & Dean, 1999). Not only does this ritual enable students to provide feedback to instructors regarding their teaching methods; but it also arms college administrators with the information needed to evaluate their faculty as well. (Adams, 1997; Hobson & Talbot, 2001; Sojka, Gupta, & Deeter-Schmelz, 2002).

Although research on student evaluations of teaching (SET) has sustained for nearly 90 years, many collegiate instructors consistently call the legitimacy and applicable use of these instruments into question (Arreola, 2007; Balam & Shannon, 2010; Beran & Rokosh, 2009a). Traditionally, learner input of instructor and course has been essential to accountability in collegiate programs (Zabaleta, 2007). During the late 1970s and early 1980s, the majority of investigations into SET concluded that such rating scales were both valid adequate (Greenwald, 1997). Nevertheless, many unanswered questions persist, and as a result, present-day researchers continue to examine both the effectiveness and legitimacy of student evaluations of teaching (Kozub, 2008).

Both adjunct and tenured faculty scrutinize how student evaluations of teaching may influence the administrative assessment of one's topical effectiveness and overall teaching ability (Ackerman, Gross, & Vigneron, 2009; Barth, 2008). SET offer collegiate administrative staff an overarching view of both the usefulness of curriculum and value of faculty (Emerson & Records, 2007). As such, said assessment scores are factored quite heavily into the substantiating of

faculty class assignments and various personal decisions, including salary increases, promotions, and granting of tenure (Beran & Rokosh, 2009b; Emerson & Records, 2007; Sprinkle, 2008). SET also aid in demonstrating institutional and instructional accountability to state, national and regional accreditation agencies as well (Ory, 2000).

As the above-mentioned personnel decisions have a great impact on the retaining of instructors and overall quality of student instruction, collegiate program directors also seek to utilize excellent assessment tools. Although, a contemporary review of literature discovered a lot of SET instruments were too drawn-out to inspire students to participate, did not have suitable validity and reliability figures, or were unsuccessful in meeting a given college's range of programmatic instrumentation needs (Emerson & Records, 2007). As SET must address the interests of several different audiences and domains of study, the objectives of SET have also been in a constant state of change as well (Ory, 2000). Correspondingly, the types of instruments used and the number of questions they contain have been adapted to the shifting administrative needs of each institution.

In fact, the most mystifying issue has been the correlation between SET and student learning. Whitworth, Price, and Randall (2002) suggested that two significant challenges impact the validity of an institution’s SET process: (1) that relevant instructional items are being measured and (2) how administrators’ will utilize this single cumulative SET score to evaluate a faculty member’s performance. As a result, the appraisal of assessment data can often be skewed, misinterpreted and misused (Sprinkle, 2008). Also, and since evaluations are not usually mandatory occurrence in the classroom, it is also difficult to adequately measure how good professors (or their courses) rate in the eyes of their pupils as a whole. Further, and most

troubling, students don't always evaluate professors on their effectiveness alone. As such, many researchers have found that students often utilize bias factors such as physical appeal and personality alongside their course based performance to formulate their teacher and course evaluations (Witt & Burdalski, 2013).

With the items above in mind and through this literature review, I examined work product related to student evaluation of teaching in collegiate undergraduate programs and more specifically, if these evaluations are expressly related to student learning outcomes. The characteristics and beliefs of faculty, student attributes and perceptions of faculty, attributes of course, the use of SET forms, their relation to good teaching, SET response rates, overarching reliability, the potential for bias, the factors of such biases, SET instrument qualities, currently employed custom instruments, development of new instruments, and new methods of utilization are also discussed in great detail.

SET Background

Not many subjects within the field of undergraduate collegiate instruction have been as well studied, documented, and long argued as the SET. Spencer and Flyr (1992) testified that the earliest instructor rating scale was issued in 1915 and nearly 90 years ago, Remmers & Brandenburg (1927) published the first article on student's attitudes toward instructors and their teaching. In these pioneering studies, Remmers (1928) confronted several of the foremost challenges within SET research, including whether the opinions of enrolled pupils coincide with those of peers and alumni (Remmers, 1928; Remmers & Brandenburg, 1927). Haskell (1997) stated that SET were first employed at the University of Wisconsin in the 1920s for informational feedback so that faculty might be more aware of student needs. Three decades

later, the University of Washington turned out to be one of the initial organizations to conduct an official evaluation and analysis of teaching and student outcomes as well (Guthrie, 1954). An extensive amount of research occurred during the 1970s, and much of the resulting findings have become well known, implemented and deemed wholly accurate (Greenwald, 1997). During this decade, which he titled the golden age of research on student evaluations, Centra (1993) also shared that a renewed interest in evaluation research advanced, including results that confirmed the legitimacy and usefulness of SET scores and reinforced their usage for formative and summative pursuits.

In 1973, approximately 30% of institutions requested that students evaluate their instructors (Wilson, 1998; Stark & Freishtat, 2014). By the mid-1990s, 80% of college campuses used some form of SET within their instructional evaluation processes (Seldin, 1999). As of the late-2000s, the University of California, Los Angeles dispersed approximately 300,000 annual SET to nearly 100 departments and programs, while concurrently providing detailed reports and analyses for said assessments to administrators as well (UCLA, Office of Instructional Development, 2012).

In many Academic Affairs Offices, SET are the most significant and often, the only gauge of instructional capability (Wilson, 1998). Seldin (1999) recounted an academic administrator as saying, "If I trust one source of data on teaching performance, I trust the students" (p. 15). Further, administrative staff believe the use of student ratings will increase a given department's ability to recognize and reward teaching excellence (Aleamoni, 1981; McKeachie, 1979), as well as help to improve instruction (Cohen, 1980; Marsh & Roche, 1993;

Menges, 1991; Overall & Marsh, 1979).

As the SET device is easy to fill out and takes little class or faculty time, it is also quite popular amongst academic administrators, and its consequential scores are the most typical method employed to assess teaching ability (Wilson, 1998, Cashin, 1999; Clayson, 2009; Seldin, 1999). The calculated median for a faculty member's SET ratings also appear equitable by its numerical rating scale and comparing this instructor rating to all-encompassing departmental averages is simple. Regrettably, many of the surveys used in faculty evaluation processes are custom made by an administrator from a given institution, which calls their structure, validity, and reliability into question. To further elucidate, Emerson & Records (2007) also found that many favored SET instruments lacked adequate reliability and validity data, were too lengthy to encourage voluntary participation or failed to provide sufficient breadth or depth to meet the college's instrumentation needs. It is also unclear if institutions should utilize SET as the exclusive source of information regarding teaching quality, and the effectiveness of assessment prompts and approaches to interpretation endure (Pounder, 2007). Faculty also believe that student evaluations are an integral component of administrative decision making (Barth, 2008; Beran & Rokosh, 2009b). Likewise, colleges have also not agreed on a unified definition of the skills and characteristics needed to demonstrate teaching excellence (Arreola, 2007, p. 98).

Purpose of SET. Providing formative comments to enrich a given faculty members instructional effectiveness is the overarching purpose of student evaluations of teaching (Campion, Mason, & Erdman, 2000; Rustagi, 1997; Thompson & Serra, 2005; Wallace & Wallace, 1998). Nonetheless, Centra (1993) shares that four benchmarks must be met for SET to

be genuinely formative. That is, instructors should acquire new knowledge from reviewing their evaluations; find value within said information; comprehend how to enhance and improve their methods of instruction, and be motivated to intrinsically or extrinsically improve their overall educational practice.

A staggering amount of colleges and universities also employ SET to validate and aid in the decision-making process regarding wage increases, course assignments, retention, promotions, and tenure (Seldin, 1993; Thompson & Serra, 2005). Thus, the systemic administrative use of SET causes faculty to influence students toward high teaching assessment scores, while students can then subsequently manipulate instructors to disperse higher grades (Crumbley, Henry, & Kratchman, 2001). As a result of the aforementioned symbiotic student-teacher influence and use of SET for staffing decisions, Hilt (2001) shared that one careless or superficial student comment on an SET form has the ability to wreak havoc on an otherwise promising professorial career. Wallace and Wallace (1998) supported said views and further suggest that beginning collegiate students do not have the ability to assess an instructor's competency in a domain that they themselves are new apprentices. As such, many researchers suggest that these types of students should complete an SET training onboarding course before having the ability to complete their end of semester assessment forms (Hilt, 2001; Kress, 2000; Wallace & Wallace, 1998).

Although some pupils may be objective while filling out an SET form, both faculty and collegiate administration must also recognize that many students' expectations are not reasonable and their weekly demands are not rational (Kress, 2000). Thus, many researchers advocate for assessment systems where students evaluate their coursework and related efforts alongside the

instructional quality of their assigned faculty member (Cohen, 1990; Kwan, 1999; Whitworth, Price, & Randall, 2002). These same scholars also postulate that a relationship exists between students' perception of their performance in a given course and their SET ratings of said assigned instructor.

The Current State of SET

As discussed above, much of the current SET research findings are inconstant and contentious. To fully understand the literature base landscape and resulting controversies, some contextual discussions are necessary. Champions of the SET process are largely comprised of those in university education administration, and who consult in the higher education field. Their optimistic outlook toward SET is well-matched with an all-inclusive atmosphere, which consists of favorable study findings, presently recognized educational philosophies, and a dispersal scheme largely centered within their academic disciplines (Aleamoni, 1999). As such, these practiced groups are so confident in their professional assumptions, that they dismiss any negative SET discoveries as “myths” (Aleamoni, 1999; Marsh & Roche, 2000). Supporters of the SET are also astounded that adverse findings continue to be unearthed in many studies as well (Franklin &Theall, 1991). As teaching evaluation inquiry is located within the higher education domain, it is assumed that most positive SET research results from those in the scholarly disciplines (Marsh & Roche, 2000).

With these items mentioned above in mind, and after almost 70 years of instructional efficacy, much of the academic research community trust that SET scores are a valid, dependable, and a worthy method of assessing instruction (Centra, 1977, 1993; Cohen, 1981; Marsh, 1984; 1987; Marsh & Dunkin, 1992; McKeachie, 1990; Murray, Rushton, & Paunonen,

1990; Ramsden, 1991; Seldin, 1984; 1993). Also, Marsh (1987) asserted that SET are the lone gauge of instructional value whose legitimacy has been systematically and meticulously recognized.

Although many scholars support the validity of SET, many opponents continue express reservations about SET usage and even outwardly dispute the results thereof (Chandler, 1978; Dowell & Neal, 1982; Goldman, 1993; Koblitz, 1990; Menefee, 1983; Miller, 1988; Powell, 1977; Rutland, 1990; Sheehan, 1975; Zoller, 1992). As an example, one well-known university's mathematics department continually refuses to partake in their annual SET process (Heller, 1984). A great deal of circumstantial evidence exists regarding instructor and staff resentment and distrust toward the use of SET scores too (Franklin & Theall, 1989).

Moreover, much of the research that supports SET validity often results from those who offer their skills for a fee or construct the rating scales outright, faculty and staff feel that their distrust is warranted (Powell, 1977). In responding to instructional researchers who claim that SET are valid, Sheets, Topping, and Hoftyzer (1995) stated that this view amounts to little more than the belief that correlation proves causation. Thus, in an environment that contains a lot of conflicting findings, it becomes relatively simple to choose research that reinforces a given individual's viewpoint (Dowell & Neal, 1982). Although, and regardless of professional differences, research must begin somewhere. That is, if any utility is to be found in the evaluation of teaching, SET must show at least a moderately strong relationship to learning (Cohen, 1981).

Academic freedom. Many opponents of SET are similarly concerned that an institution's questionnaire results may be used as trappings to exert managerial dominance and political

control over faculty, and that the obligation of semester-ending evaluation endangers academic freedom (Foucault, 1980). The academic freedom of untenured faculty is decreed both sensitive and vulnerable to SET, and the dissection of their results can be employed to rationalize directorial decisions (Wicks, 2004). To further elucidate, SET places instructors under scrutiny, dampens academic freedom and escalates administrative oversight by delivering an instrument of control that impacts curricula, lecture content, instructional practice and grading criteria (Haskell, 1997a, 1997b).

The challenges of measuring teaching effectiveness. The characteristics of effective instructors must be acknowledged and subsequently stated in a manner in which they can be both explained and assessed (Cassidy, 1990). Although most academics agree that there is a need for instructional evaluation, they do not concur on whether or not the present-day instruments utilized have a valid application for this purpose (Marsh, 2007). Both the SET research and their related assessment tools are hampered by the same challenge, i.e., there is no consensus on (or commonly accepted definition of) what "good" teaching is, nor has an entirely agreeable standard of teaching effectiveness been established (Adams, 1997). Further, and since disparate courses also do not follow a standard method of dispersing assignments, administering exams or determining a grade, quantifying learning can be difficult. Grades are not perceived as one's full measure of learning and can be a poor gauge of student progress (Beleche, Fairris, & Marks, 2012).

Feldman (1986) stated that a successful teacher is proficient at constructing circumstances favorable to learning, and those that do and do not support the SET widely agree that students will learn more from quality individuals and good teaching. Cohen (1981) affirmed,

"Even though there is a lack of unanimity on a definition of good teaching, most researchers in this area agree that student learning is the most important criterion of teaching effectiveness" (p. 283).

To speak to the multi-dimensionality of instructional objectives and accountabilities, Hobson & Talbot (2001) further shared that focusing on student learning alone has the potential to exacerbate the struggle of defining teaching effectiveness, because learning assessment embodies an equally broad criterion, which includes an individual's development, self-discipline, trajectory toward mastery, and career progress. Further, Marsh (1987) identified nine instructional elements that are related to effectual teaching: the usefulness of information, eagerness to share information, clarity and organizational skills, group interaction, rapport with pupils, scope of topical coverage, consistency in grading and crafting examinations, relevant assignments and readings, and overall difficulty of workload.

Although a large amount of research supports Marsh's multidimensional views above, a contemporary movement toward outcome assessment further promotes measuring student learning in conjunction with job placement too (Ory, 2000). Scriven (1981) further observed, "The best teaching is not that which produces the most learning, but that produces the best outcome" (p. 248). This career-centric and outcome-based methodology seeks to replace student rating systems and their evaluation-based administrative decisions, which critics of SET believe are too subjective. Much of the subjectivity in the SET process results from students evaluating zones of instructional efficiency that they are incapable of inspecting, which have been identified as the ambitions, substance, and organization of course design; the techniques and resources used in delivery; and the appraisal of student work, including grading criteria and practices (Cohen,

1981; Cashin, 1988; McKeachie, 1997; 2002; Bain, 2011). However, if an evaluation process is effective, then there should be a relationship between student learning and the evaluations that students fill out for their classes and instructors (Cohen, 1981).

Although, each of the above-mentioned limited views of student teaching assessment also restricts a given academics command of what is being measured by SET, what can be deduced from SET, and how conclusions from various reports can be comprehended within a common structure. Thus, Marsh (1987) encourages an intertwined construct validity approach in which SET can be positively related to a diverse assortment of effective instructional benchmarks (i.e., good teaching), and where specific scoring elements are conceived to be most associated with variables to which they are most reasonably and academically related. Additional measures could also include transformations in academic student behaviors, faculty-led self-evaluations, assessments fulfilled by colleagues and management, and in-class observations viewed by trained personnel. This intertwined construct validity approach of SET scholarship is now acknowledged by a wide variety of institutions and scholars (e.g., Cashin, 1988; Howard, Conway, & Maxwell, 1985).

Although a positive step, the approach as mentioned earlier outlined by Marsh assumes that each SET domain is reliably measured and that it authentically reflects efficient instruction. That is, if said alternate benchmarks of instructional achievement are not consistent and usable, then they should not be employed for additional purposes within an institution's SET research, academic policy creation, feedback to faculty, or personnel pronouncements (Abrami, d'Apollonia, & Cohen, 1990; Marsh & Dunkin, 1992, 1997; Marsh & Roche, 1993). With these items in mind, and when one examines the literature much more closely, there are numerous

additional topics that emerge as relevant when attempting to evaluate the learning/SET association. These are detailed in the forthcoming sections.

Faculty characteristics. Numerous characteristics of faculty are associated with favorable SET scores. Good evaluations usually result from those individuals who are organized, deliver accurate and uncluttered lectures and communicate in a clear manner (Donnon, Delver, & Beran, 2010). Teachers perceived as passionate, friendly, cooperative, and impartial, were rated as more effectual instructors than those who did not have these qualities attributed to them (Barth, 2008; Hills, Naegle, & Bartkus, 2009). Further, the rigor of a course and personal unapproachability alongside a lack of organization, clarity in teaching, and poor time management, were common faculty deficiencies associated with subpar SET scores (Barth, 2008). Another premise found in much of the literature reviewed includes the "leniency hypothesis" or otherwise titled the paradox of rigor. This theory assumes that instructors who employ a lenient grading scheme may gather more favorable SET and as a result -- achieve superior overall evaluation scores (Gump, 2007; Heckert, Latier, Ringwald-Burton, & Drazen, 2006).

Level of instructor and related experience. When the SET scores of tenured faculty are compared to teaching assistants, faculty receive higher marks (Centra & Creech, 1976; Marsh & Dunkin, 1992). Centra (1978) also shared that faculty in the later years of their career consistently earn higher scores than first-year instructors as well. Higher rankings on evaluation instruments were also granted to those with more teaching experience or expertise within a given area of study (Nowell, Gale, & Handley, 2010). Adjunct faculty received higher marks from new enrollees, but upperclassmen seem to favor tenured instructors (Stark & Freishtat, 2014).

Notwithstanding the studies that contained teaching assistants, Feldman (1983) found no significant relationship between faculty rank and SET ratings. Although, when he reviewed research which uncovered a significant (though weak) relationship between faculty rank and SET, the more elevated instructors received more favorable SET scores. Further, when Feldman (1983) reviewed studies in which a substantial association was identified, he discovered that faculty of advanced age and teaching experience earned lower ratings. As most studies have been cross-sectional (i.e., rather than longitudinal), Centra (1993) believed that the effect of rank and experience on SET are incomplete and require further investigation. Finally, no meaningful correlation exists between SET and an instructors age or experience level either (Feldman, 1983).

Faculty reputation. Although much of the SET literature does not discuss a faculty member's professional reputation, Perry, Niemi, and Jones (1974) shared that one U.S. University's students' prior expectations of a notable professor's teaching performance influenced SET scores. Supporting this outlook, students who utilized instructor reputation to select and enroll in a course also furnished higher faculty SET scores than their associated classmates (Leventhal, Abrami, & Perry, 1976). Also, highly animated faculty with an undesirable reputation usually receive a lower SET score than those who are highly animated alongside positive reputations; but most interestingly -- those demonstrating a lesser amount of self-expression were not meaningfully affected by the variable of reputation (Perry, Abrami, Leventhal, & Check, 1979).

Faculty research. There are two differing opinions regarding faculty research productivity. On the one hand, some intellectuals believe that further scholarly research helps an

individual to stay current in their chosen field, thereby improving one's teaching efficiency. Opposing academics feel that those faculty who spend additional moments on scholastic endeavors have fewer opportunities to devote time to their instructional practice. Perhaps most interestingly, research demonstrates that no significant connection exists or a very weak positive association exists between research output and SET (Aleamoni, 1987; Centra, 1983; Feldman, 1986; Marsh, 1979, 1987; Marsh & Dunkin, 1992).

In contrast, Allen (1995) identified a minor but noteworthy positive correlation between faculty research and SET. Centra (1983) also uncovered a modest positive correlation between faculty research and SET in the social sciences alone. Despite these findings, universities still review a faculty member's research and publication track record to offer tenure and promotion as well as evaluate their teaching (Seldin, 1984). This tradition has been extensively condemned (Aubrecht, 1984; Centra, 1993).

Gender of instructor. Conversations regarding the associated impact of teacher gender on SET are reasonably varied. Several researchers believe that SET are predisposed to negatively impact female faculty (Basow, 1994; Basow & Silberg, 1987; Kaschak, 1978; Koblitz, 1990). Also and to aid female faculty in sidestepping getting lower ratings than their male colleagues, they are encouraged to behave in conventional feminine manners (Bennett, 1982). As such, female faculty who embrace a more rigorous instructional approach to workload normally receive lower SET scores (Koblitz, 1990). Although, studies by Tatro (1995) report female instructors obtain significantly higher ratings than male instructors on semester-ending SET.

Feldman (1992, 1993) also examined the prevailing SET research related to both male and female instructors, which featured both laboratory and classroom locales. In the laboratory

venues, Feldman (1992) reported that many of the studies examined did not display a resulting difference in SET between male and female teachers. In the marginal number of readings in which variances were uncovered, female faculty received lower ratings than males. In later research, Feldman (1993) testified that many classroom studies did not display significant differences between faculty genders. Nevertheless, Feldman (1993) also shared that in the limited number of studies where classroom variances were revealed, female instructors received marginally higher SET scores than males. Students were also inclined to score same-gender faculty marginally higher than opposite-gender instructors as well.

Minority race of faculty. It is unclear if a systematic racial bias in SET exists (Centra, 1993). Though, Rubin and Rubin (2011) believe that non-native speaking faculty comprised of various races score lower on SET. A report by Buck and Tiene (1989) found a significant interaction between attractiveness, gender, and race; that is, teachers with an unfamiliar race were rated slightly higher if they were both female and good-looking. Additionally, students' perceptions of a non-native speaking faculty member's teaching ability were touched by opinions of outward appearance (Rubin & Rubin, 2011).

The Dr. Fox study. In the Naftulin et al., (1973) study entitled 'Dr. Fox', a professional actor, referred to as 'Dr. Fox', administered an enthusiastic and charismatic lecture devoid of instruction that persuaded a class into giving undeservedly high SET scores. Not only was this research harshly condemned; but the topic of seductiveness was also reanalyzed by several subsequent studies (Marsh, 1987) with the 'Dr. Fox' receiving little further literature support (Perry, 1990). Also, an academic literature review by Abrami and Mizener (1983) discovered that SET were much less sensitive to lecture content that overall instructor self-expression.

In later research, Abrami (1989) shared that faculty expressiveness impacts SET beyond student learning, which represents a bias. Additionally, when students were not presented with encouragement to participate, instructor self-expression had a larger impact on SET scores than lecture content; however, when encouragement was shared, expressiveness was far less significant and no resulting 'Dr. Fox' effect was found (Marsh & Ware, 1982).

Faculty perceptions of the SET process. When reviewing studies that investigate faculty perceptions of the SET process, there appears to be a general divide amongst faculty too. Several studies documented positive faculty views of the evaluation process and most stated that the SET aids them in making meaningful improvements to their teaching practice (Balam & Shannon, 2010; Beran & Rokosh, 2009b; Beran & Violato, 2006). In short, if faculty view their student ratings and the overall process of evaluation as positive, they are also more likely to state that the resulting scores are useful (Beran & Rokosh, 2009a; Beran & Rokosh, 2009b). However, opposing studies indicate that many instructors are suspicious of the validity of the evaluation instrument, as well as the time, effort and attention that students attribute to completing their SET. In addition, several faculty view SET scores as a measure of a given instructor's popularity (Chandler, 1978; Clayson & Sheffet, 2006) and worse yet, as an occasion for students to retaliate against a demanding teacher (Wright, 2006). As expected, numerous faculty members further believe that SET scores are not legitimate, reliable and that such feedback does nothing to help boost their teaching performance (Balam & Shannon, 2010; Beran & Rokosh, 2009).

Some faculty even believe that the period for administering the evaluation, i.e., both the time of day and cycle during the semester may also sway the results of student opinion (McNulty, Gruener, Chandrasekhar, Espiritu, Hoyt, & Ensminger, 2010). Countless faculty

members also do not believe that students take in-class evaluations seriously (El Hassan, 2009). Further, nearly all instructors believe that students are not experts in classroom instruction or the overarching concept of pedagogy and as a result, they can only judge their reaction to a course (Ackerman, Gross, & Vigneron, 2009). However, when faculty are well informed about their given institution's reasons for assessment, much of their anxiety disperses and eagerness to learn from undergraduate feedback increases (Sojka, Gupta, & Deeter-Schmetz, 2002; Hativa, 1995; Gallagher, 2000; Bain, 2011).

Student characteristics and views. While no reliable association exists amongst student characteristics and SET, many studies agree that students are perfectly situated to give feedback regarding their experience in a course (Abrami & Mizener, 1983). Fraser and Treagust (1986) agree that students are unavoidably in the classroom and as a result, perfectly placed to report on their instructor. Further, a student's attendance does not alter the scope of what occurs in the classroom or generate a scholastic Heisenberg effect (Page, 1974). Thus, students are situated to report on factors that guide teaching effectiveness, such as a faculty member's linguistic clarity, writing legibility, and possibly the instructor's availability during office hours and the like (Stark & Freishtat, 2014).

Numerous studies also suggest that students who have prior knowledge or curiosity in the subject area preceding a given course usually provide higher SET scores (Feldman, 1977, p. 236; Marsh & Cooper, 1981; Prave & Baril, 1993). They can also mention whether they feel more inspired about the subject matter (upon conclusion of the course) and if said class motivated them to take a related or follow-up serial course. As clarity may be confounded with the difficulty of class material, clearness is also harder to ascertain and interpret as well.

Also, several student characteristics were linked to elevated student appraisals of faculty. Students who anticipated earning a high grade were more likely to deliver higher ratings (Barth, 2008; Beran & Rokosh, 2009b; Beran & Violato, 2006; Beran & Violato, 2009; Olshavsky & Spreng, 1995; Nowell et al., 2010; Serdyukova, Tatum, & Serdyukov, 2010). In addition, students that had high expectations and favorable class experiences provided higher SET scores than those with low expectations and favorable class experiences (Koermer & Petelle, 1991). Anticipated grades frequently correlate with positive ratings of professors, and rigor is often negatively related or believed to be negatively related to SET (Clayson, 2009; Clayson & Haley, 2011; Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Sojka, Gupta, & Deeter-Schmelz, 2002). Regrettably, and in addition to expected marks, reports of student satisfaction may also be based on an instructor's complementary personality (Betoret, 2007), and social factors (Blackhart, Peruche, DeWall, & Joiner, 2006) rather than on actual learning outcomes (Marks, 2000; Simpson & Siguaw, 2000).

Furthermore, students who regularly attended and participated in class provided the highest marks to instructors (Beran & Violato, 2006; Hills, Naegle, & Bartkus, 2009). However, only a small amount of the variance amongst SET ratings can be attributed to students who achieve high grades. To further clarify, the best and worst students report the most favorably on SET in a given course (Serdyukova, Tatum, & Serdyukov, 2010). Often, students stress the importance of a quality in-class experience but concurrently rank their participation in the course as less meaningful. As such, Kress (2000) suggested that students should concurrently assess their progress within a course and how that progress relates to the performance of their instructor. Finally, and perhaps most interesting, physical attractiveness also matters; i.e., faculty

evaluations can be predicted from an individual's reaction to 30 seconds of silent video of the teacher (Ambady & Rosenthal, 1993).

Life experiences, age, and gender may also sway SET. First and second-year students perceived their instructors more critically than those in years three and four (Clayson, 2009). Ease of workload, application of course materials and grading were the main concerns in freshman and sophomore classes, whereas juniors and seniors were more likely to place emphasis and enjoyment on in-class discussions and activities (Hills et al., 2009). All students placed value on transparency in their classes alongside the modern course related relevance to success in the real world.

Although many students do not perceive the purpose and utilization of SET in the same manner as faculty, they do believe that they can make accurate evaluations of faculty teaching (Balam & Shannon, 2010). In one such study by El Hassan (2009), a quarter of students conceded that SET are inconsequential and that they only partook because it was mandated. Nearly one-third of these same students stated that earning a five on the faculty rating scale was unreachable and more interestingly -- when they were either noncommittal or indifferent in the SET process (or about their instructor), just over one-half of them marked a rating of 3 on a 5-point scale. Another third of the studied students testified that they gave their faculty member a higher or lower evaluation than the instructor deserved. While filling out the SET, nearly one-fifth stated that they had penned something false (Clayson & Haley, 2011).

Students are also more eager to contribute and volunteer significant comments when they trust and can tell that their feedback is being contemplated and implemented by their faculty and the college (Chen, & Hoshower, 2003). However, the clear majority of learners do not believe or

recognize that their input is being utilized. Additional findings indicate that students place the greatest importance on SET for formative purposes, but studies also imply that students believe their feedback should be measured for summative purposes as well (Clayson & Haley, 2011). Regarding overarching assessment content, students would prefer that SET instruments feature more specific items linked to teaching effectiveness too (Sojka & Deeter-Schmetz, 2002; Chen & Hoshower, 2003).

Faculty evaluations can also be adversely affected by student behavior. As an example, and even though the instructions were repeated three times, Clayson and Haley (2011) noted in their studies that just over 10% of students did not follow the prompts correctly on at least one part of their survey form. As such, one may surmise that errors are occurring during the SET process, which may lead to wholly inaccurate results.

Student gender. Studies that examine the connection between student gender and ratings are mixed in their findings (Aleamoni, 1987; Aleamoni & Hexner, 1980). Numerous reports declare that SET scores by female and male pupils are identical, whereas others have reported the opposite conclusion. Amongst studies that did display noteworthy affiliations between student gender and SET, most discovered that males provided lower scores than females (Feldman, 1977, Tatro, 1995). However, in a contrasting study, males provided marginally higher scores than female students (Koushki & Kuhn, 1982). As described in previous sections, additional findings display a susceptibility for learners to score same-gender faculty somewhat higher than opposite-gender faculty (Feldman, 1993; Centra, 1993).

Emotional state of students. In a singular research study, the emotional state of students was associated with SET (Small, Hollenbeck, & Haley, 1982). To further clarify, the more

argumentative, restless and disheartened students felt at the end of their semester, the lower they scored their instructors on SET. Although this may be an overarching and severe risk to the validity of end-of-semester administered SET, Wachtel (1998) stated that the results above have not been supported by recent studies.

Administration of SET. One should not undermine a suitable SET form with unacceptable organizational processes, directives or scheduling (Seldin, 1993). Thus, said administrative features are contemplated in the following segments below.

Timing of evaluation. No matter if a SET is dispersed in the midst of a course or at the conclusion of the final exam, the time at which a course's SET is administered does not affect student ratings (Feldman, 1979). Frey (1976) also exposed that student scores gathered during the last seven days of a term were not considerably dissimilar from SET gathered during the first seven days of a subsequent term. Further supporting this finding, Marsh and Overall (1980) discovered that mid-semester scores were decidedly related to semester-end ratings, but Aleamoni (1981) declared that the results of a SET that is distributed before or after a final exam may be skewed. As such, Braskamp and Ory (1994) advise academic administrators to disperse SET during the last fourteen days of a given semester, without interfering with the last day of class. When SET are to be employed in a summative manner, Seldin (1989) agreed with the previous recommendation, but further, suggests that the 30 to 45-day mark of a given course is ideal for dispersing questionnaires for formative purposes. Although, and as there are discrepancies within the schedules discussed above, L'hommedieu, Menges, and Brinko (1990) call for additional research into the effect of timing on SET.

Anonymity of participants. When students identify themselves on a SET, they tend to furnish higher faculty ratings compared to those individuals who stay unidentified (Feldman, 1979; Blunt, 1991). However, other circumstances may intermingle with namelessness, including (but not necessarily limited to) if the SET rating is given before or after students receive their final grades, if it is said that the scores will be utilized solely for research purposes and whether the students believe that an instructor may be able to review their responses (Abrami, Perry, & Leventhal, 1982). Many students are also unsure if collegiate administration can fully guarantee that SET responses are both anonymous and confidential (Blunt, 1991). A multitude of studies also support that each student response must remain wholly anonymous (Braskamp & Ory, 1994; Centra, 1993; McCallum, 1984).

Effects of sharing SET data with students. Although most colleges prefer to keep SET responses confidential, both the University of Idaho and the University of Wisconsin determined that students had the right to view SET results (Haskell, 1997). As such, students who do not have access to SET scores rate these assessments as more important to their course selection than those who can view them (Wilhelm & Comegys, 2004). Thus, and if all other attributes are equal, students are twice as likely to choose a faculty member who has received excellent SET scores versus those who received average or low ratings (Wilhelm, 2004). Although, if students believe they will learn a great deal from a given class they are willing to choose a substandard faculty member who has received poor SET scores (Wilhelm & Comegys, 2004). Outcomes are varied on whether obtaining SET scores in advance of enrolling in a course influences the future ratings of an instructor. Many research reports have shown that students who obtain evidence

that a faculty member was scored highly - will score that same instructor highly, and vice versa (Perry, Niemi, Jones, 1974; Haskell, 1997).

Instructor presence. When the faculty member being assessed is present in the classroom during the evaluation process, student ratings are often elevated (Feldman, 1979). As a result, many scholars suggest utilizing an outside firm (i.e., one that is not related to the institution in any way) to collect SET responses and subsequently protect student anonymity (Braskamp & Ory, 1994; Eble, 1970; Scriven, 1981). Even if he or she is not the individual who collects the SET forms, faculty who initially distribute said surveys might also inadvertently impact student responses (Pulich, 1984). As a result, Pulich (1984) further recommended that an independent assessor circulate and gather each SET form, as well as answer student questions or clarify procedural items during the evaluation event. In short, the SET process should be devoid of faculty participation or attendance.

Purpose of evaluation. If the stated intent of a SET is for tenure or promotion, student ratings are usually higher than those that are employed for academic purposes alone (Aleamoni & Hexner, 1980; Centra, 1976; Feldman, 1979). However, Frankhouser (1984) determined that the stated purpose of a SET had no substantial influence on student scores. Although, it is suggested that students be informed if SET responses are to be utilized for employment decisions, including class assignment, promotion and the like (Braskamp & Ory, 1994).

Attributes of course. Notwithstanding the ability of the instructor, the attributes of a course may have an effect on SET (Marsh, 1987). These include the electivity, meeting time, level, size, subject, featured workload, and classroom environment of a course, as well as its related departmental considerations.

Elective Courses. Required courses had lower student ratings than electives (Feldman, 2007). Elective courses are described as classes that have a high percentage of students who are taking it outside of their core area of study (Feldman, 1978). As there is typically lower student interest in required versus elective courses, a modest positive correlation was established between non-required classes and student SET scores (Brandenburg, Slinde, & Batista, 1977; Feldman, 1978; McKeachie, 1979; Scherr & Scherr, 1990).

Class meeting time. Although limited research studies have been conducted on this topic, most experts believe that a relationship does not exist between SET scores and the time of day a given course congregates (Aleamoni, 1987; Centra, 1993; Feldman, 1978). However, one such study discovered that extremely early morning classes, classes after lunch and late afternoon classes receive the lowest SET ratings (Koushki and Kuhn, 1982). Thus, other contextual variables including gender, expected earned grade, pursued degree, and year in college had less of an impact on SET ratings than a class' meeting time (McKeachie, 1979).

Level of course. A great deal of research also suggests that higher level classes receive superior SET scores (Feldman, 1978; Marsh, 1987). The relationship between class level and SET scores is also weakened when other contextual items such as the size of class, anticipated grade earned, and elective or required course type are controlled for as well. As such, the effect of class hierarchy on SET ratings may be direct, indirect, or equally attributed to both.

Further, much of the current writings have also largely overlooked the age of students within a course level as well. Thus, the variance in student age (or associated maturity thereof) at the time the assessments are dispensed could be a larger source of influence on SET scores than the course characteristics themselves (Centra, 1993; Feldman, 1978).

Class size. Smaller classes usually receive higher SET scores (Feldman, 1978; Franklin & Theall, 1991; McKeachie, 1990). Particular dimensions of effectual instruction are also influenced by class size, such as faculty rapport with students and class participation (Marsh 1987; Marsh & Dunkin, 1992). Although, Abrami (1989) in his criticism of Marsh's (1987) study counters that this view must not be employed to support the legitimacy of SET scores, and further demonstrates that rapport and participation, being sensitive to class size, are dimensions that should not be utilized in summative choices. In addition, when instructors perceived their class size to be too large to present course material sufficiently, they received lower ratings than other faculty members (Scott, 1977). Thus, an instructor's perceptions about class size may impact teaching performance and the resulting SET class ratings (Feldman, 1978).

Additional studies also postulate the connection between student SET scores and class size is not horizontal, but instead curved, with both large and small classes obtaining better scores than mid-sized occupancies (Centra & Creech, 1976; Feldman, 1978, 1984; Koushki & Kuhn, 1982). If more researchers had been aware of this U-shaped theory, they might also have also uncovered a curvilinear relationship in their studies as well (Feldman, 1978). Somewhat surprisingly, elevated SET ratings may occur in large classes for the following reasons: due to the possible increased pressure of teaching larger classes, instructors may prepare in a more thorough manner or adjust their teachings to accommodate the meeting size, including additional visual aids and robust study materials; prominent instructors may entice a greater number of students to enroll in a given course, thereby producing a larger class size; program chairs or departmental leads may appoint their most proficient instructors to sizeable classes; and courses

that exceed one hundred enrollees often feature multiple teaching assistants alongside small discussion groups.

Subject of class. The subject matter presented in a given class influences student SET scores (Ramsden, 1991), with science and arithmetic studies placing among the lowest ratings in collegiate studies (Cashin, 1990, 1992; Centra & Creech, 1976; Feldman, 1978). Centra (1993) shares that the math and science domains are adversely affected due to their classes being faster paced, less student-oriented and that tenured faculty must also allocate a great deal of time into seeking grants and fulfilling research for the institution (i.e., than faculty in other disciplines). In addition, instructors and students in the science fields often have unusual perceptions of workload and pace of course, which can also negatively impact SET scores as well. However, if teaching is less effective in topics that consistently receive lower SET scores, then the topic of a class would not produce rating bias; but if topics demanding quantifiable cognitive abilities are scored lower because students are less capable in said proficiencies, then a bias to scores would result (Cashin, 1988).

As large disparities exist between disciplines, comparisons in SET scores amongst departments or their faculty should also not be made (Ramsden, 1991). Furthermore, a substandard instructor who presents attention-grabbing curricula is consistently rated higher on SET than an excellent instructor exhibiting uninteresting material, which further discourages the comparing of SET scores amongst unconnected subjects (Perry, Niemi, & Jones, 1974).

Course workload. Notwithstanding the ability of the instructor, the attributes of a course may have an effect on SET. Lower SET scores were attributed to courses with high workloads (Guder & Malliaris, 2010). Although, Barth (2008) stated that a teacher could reverse these

effects if additional assistance and support is provided to students via additional class time, office hours and the like. Although, Marsh (1987) found that challenging classes with increased workloads received more favorable scores, thereby rejecting these items as a potential prejudice to SET. Dudley and Shawver (1991) also cited a study in which an advertising course without relevant coursework was vastly enriched by the introduction of applicable homework; however, this outcome may not be relevant to all domains. Ryan, Anderson, and Birchler, (1980) also contradicted these items by sharing that compulsory SET frequently lead faculty to reduce the number of class assignments and craft easier exams. A course's difficulty and workload may also not be uniform. To further clarify, Franklin and Theall (1991) divided difficulty and workload into two separate variables and as a result of this method learned that difficulty (not workload), had a minor positive relationship to SET scores.

As discussed in the aforementioned paragraphs, a faculty member's perception of course, pace and workload may noticeably contrast from that of students. Further, the assumption that classes with increased relevant coursework result in higher SET scores is not found in the literature related to arithmetic and science studies (Cashin, 1990). Therefore, course pace or a student's perceptions thereof may result in lower SET ratings.

Class environment. Other items not in a faculty member's purview, such as a classroom's atmosphere also have an impact on SET ratings. Examining a study in which faculty taught identical classes using the same syllabi, exams, assignments, and PowerPoint slides, students who were enrolled in more modern or improved space remarked that they were learning a great deal and that their instructor was very organized, i.e., compared to students in rooms with a good amount of noise, less-than-perfect lighting and uncomfortable seating (Hill & Epps,

2010). A great deal of these class attributes live outside of a given faculty members jurisdiction (Nowell et al., 2010) yet the class location can have a significant effect on student evaluation of teaching results. In contrast, a lone study did not find that a classroom's location directly influenced SET; but rather that the course characteristics were mediated through student engagement (Beran & Violato, 2009).

Departmental considerations. While Cranton and Smith (1986) agree with the overarching studies in the sections mentioned above, they also found that the effect of course attributes on SET scores varied significantly between collegiate departments. In select branches, characteristics of course did not have a noteworthy effect on SET scores, while in other departments student ratings contrasted with what had been predicted. As such, the researchers determined that campus-wide patterns for teaching efficacy cannot be established. As the correlation between course attributes and SET scores is not equal amongst dissimilar departments, Cranton and Smith (1986) lobby for additional research to be conducted on the effects of course characteristics in singular domains.

Online versus paper SET. With the development of the internet and campus-wide access, online SET are becoming more widespread in higher education (Thorpe, 2002). In a Hmieleski and Champagne (2000) study of the 200 most tech-savvy campuses, only two reported dispersing college-wide SET online. No schools reported employing mobile devices to complete SET (Thorpe, 2002).

Anderson, Cain, and Bird (2005) reviewed the literature regarding online dispersal of SET and subsequently outlined the advantages and disadvantages thereof. Advantages include fast response times; lower costs to disperse, collect and measure data; allows an infinite amount

of outside of class time to complete; is less susceptible to instructor influence, and allows students multiple opportunities to craft personal responses. Disadvantages include: one must have access to both a device and the internet; produces lower response rates, and it is perceived as less accurate by veteran faculty.

Additional scholars have also compared these more modern online methods of evaluation to traditional paper SET forms. When distributed through learning management systems such as Blackboard, Canvas or Brightspace, students, faculty, and staff generally view online evaluations more positively than paper evaluations (Anderson, Cain, & Bird, 2005). Online forms may be favored for many reasons including asynchronous and mobile device participation, ease of modifying responses and ability to insert individual comments. Still, online asynchronous SET questionnaires lead to harsh or compassionate assessments, alongside wholly inaccurate faculty ratings (Nowell, Gale, & Handley, 2010)

In addition to traditional university-administered questionnaires, web-based self-governing sites such as RateMyProfessor.com (RMP) allow students to share their instructor feedback publicly, and as a result, most students believe that this type of SET is trustworthy (Brown, Baillie, & Fraser, 2009). Further, and when compared with conventional paper course evaluations, additional studies have also concluded that RMP ratings are valid (Brown, Baillie, & Fraser, 2009; Sonntag, Bassett, & Snyder, 2009). However, in likening RMP to college dispersed evaluations, these reports only observed the perception of a course's difficulty alongside the overall quality of a course (Brown, Baillie, & Fraser, 2009; Sonntag, Bassett, & Snyder, 2009).

Demonstrative Research. In 1998 at Rutgers College of Pharmacy, Woodward conducted one of the earliest studies that compared paper SET with digitally dispersed versions thereof.

While focusing on a single fall semester 3-credit course, students were arbitrarily separated into two factions – i.e., those that would complete paper SETs and those that would complete online versions of the same form. Comparing the results from both 1996 and 1997, the demographics of the two classes and resulting student factions were equivalent, and the instructor assessment scores and student responses to open-ended inquiries were similar as well.

Layne (1999) randomly assigned traditional paper or online course evaluations to nearly 2,500 computer literate students at a sizeable southeastern institution. As the same assessment form was employed in both SET dispersal methods, the researcher found that pupils were more willing to assess their instructor when the forms were administered in a traditional class setting, but the average instructor score did not vary between the two delivery schemes. Although, and somewhat startling, students who fulfilled their SET digitally provided more commentaries regarding both their instructor and course than those individuals who took part in the traditional in-class paper assessment methodology.

Kasiar, Schroeder, and Holstad (2001) furthered the comparisons between paper and online SET and as a result, employed 169 students in an undergraduate pharmacotherapy course, taught by multiple professors. Utilizing identical SET surveys in both formats, 50 learners were chosen at random to participate online, while the remaining 119 students partook in the more established paper process of evaluation. Upon completion of the study, the findings were as follows: students participating in the digital SET offered more comments and typed nearly 7 times more words than those students who filled out paper forms; students took 10 minutes (or less) to complete the digital SET vs. nearly 30 minutes for the paper assessment, and the

workload of collegiate administrative staff was significantly reduced. To further clarify, clerical employees spent roughly 30 hours to collect the marks and commentaries from the school's paper SET and only 1 hour to download this same information from the digital version thereof. Both the authors and institution concluded that the decreasing staff and student workloads alongside the prompt reporting of SET results were advantageous and as a result, they planned to employ more digital SET throughout the curricula.

Responses rates. Many students do not participate in collegiate SET surveys, and therefore, the resulting response rate will be below 100%. The lower the response rate, the less demonstrative the responses may be overall. For example, frustration, disappointment, and anger motivate people to action, and to complete a SET more than satisfaction does (Stark & Freishtat, 2014). Thus, the response rates themselves say little about the instructional value or the quality of instruction. This data must not be considered to represent a whole cohort or class, i.e., the average taken of small class size samples are more susceptible to chance or bias - than the averages taken from larger samples. As such (and even if the response rate is 100%.), small classes may have radically different SET results than evaluation administered in larger classes (Feldman, 1984; Marsh, 1987). Students in classes with low enrollment might feel that their secrecy is more questionable, which could also reduce their willingness to respond truthfully or to participate at all (Perry & Smart, 2006).

Comparison of online and paper response rates. The migration from paper to online or digital SET has also had an adverse effect on response rates, which saw many institutions' amounts drop by nearly 25%. Due to these new polling methods, larger class sizes did not perform well, while smaller class sizes had better digital response rates overall (Guder &

Malliaris, 2010). With the previous items in mind, the overarching findings related to online response rates and SET are divided. Nowell et al. (2010) found that online SET completion had a negative effect on SET scores (as compared to paper scores), whereas Guder and Malliaris (2010) found no significant change, i.e., other than students made a greater number and lengthier comments in their digital systems. However, many studies concur that digital out-of-class SET surveys lead to severely coarse or lenient evaluations, alongside lower response rates and diminished faculty ratings (Nowell et al., 2010; Serdyukova, Tatum, & Serdyukov, 2010).

Richardson (2005) shared that little information exists regarding the response rates of digital SET, or whether diverse methods of dispersal produce comparable results or patterns. However, upon closer review of the current literature, a good deal of information is widely accessible regarding the differences in response rates between digital and traditional paper SET. For example, of the eight studies outlined in Table 1 (below), the clear majority of digital SET response rates were lower than those employing traditional paper distribution. That is, online methods averaged a 33% rate of response compared to the 56% attained in paper dispersals.

Table 1
Paper-based vs. Online SET Response Rates

	Paper-based response rate (%)	Online response rate (%)	Difference (%)
Cook et al. (2000)	55.6	–	−16
Baruch (1999)	–	39.6	
Dommeyer et al. (2004)	75	43	−32
Ballantyne (2005)	55	47	−8
Ogier (2005)	65	30	−35
Nair et al. (2005)	56	31	−31
Griffith University (2005)	57	20	−37
Sweep (2006)	56	23	−33
Watt et al. (2002)	32.6	33.3	<1
Overall	56	33	−23

However, there are three exceptions to the prevailing viewpoint discussed above. The first of which is summarized in the research conducted by Watt, Simpson, Mckillop, and Nunn (2002), who found that online SET response rates reached nearly 33% while paper versions thereof totaled 33.3%. As this data is inconsistent with the other research shared in Table 1, one must look more closely into Watt et al.'s (2002) paper-based low response rate information. In fact, the courses assessed on paper forms within said research were taught in distance learning programs and as a result, said traditional handwritten forms were not dispersed in a classroom (i.e., face-to-face) setting. As such, one could surmise that the dispersal of SET in a face-to-face environment results in higher response rates. Although, the improvement of digital response rates if administered in a classroom face-to-face environment remains unknown.

The research from Dommeyer, Baum, Hanna, and Chapman (2004) contains the second exemption to the information distributed in Table 1. To further clarify, in 14 of 16 instances and significantly so in 10, the response rates for paper-based SET were higher than those SET conducted online. When prospective respondents were presented with an incentive of a quarter of 1% increase in total course grade, response rates were not significantly affected or different. Although, when a grade incentive was tallied for those pupils who would not have otherwise participated in SET, both paper and online SET response rates were nearly identical and very high, i.e., approximately 87% each. In general, Dommeyer et al. shared that traditional paper SET received a response rate of 75%, whereas digitally distributed versions only reached 43%. Finally, Guder, and Malliaris (2013) share a limited study in which paper evaluations reach 70–80% participation, while online evaluations enlist 50–60% of student responses.

Recommendations for improving paper SET response rates. Much of the current literature suggests that most administrators employ three methods to improve end-of-semester SET response rates. The first and most effective means is to make the evaluation part of the course requirements – i.e., dispersing a SET around the time of midterm exams and subsequently sharing the results and your plan of action to students based on their feedback (Marsh, 1987). Sending reminder notices and offering small incentives to students are also helpful in securing higher rates of student response (Beran & Rokosh, 2009). For example, at Columbia University, three weekly email reminders are sent to students accounts until the course SET is administered, while other faculty members concurrently offered either a one-half of one percent grade increase or prizes for participation as well.

Methods to improve online SET response rates. The three most prevalent methods for boosting online survey response rates are repetitive reminder emails to student non-respondents; repetitive reminder emails to academic SET owners; and securing various incentives for student responses including prizes and the like (Nulty, 2008). In addition, two credible resources share succinct, somewhat similar and interrelated advice regarding methods that may have the ability to raise institutional SET rates of response as well.

The first, Zúñiga (2004) from the US Teaching and Learning with Technology/Flashlight Group, outlined their faction's seven best exercises for increasing digitally dispersed SET response rates. These are: send the survey directly to a given student and include a URL that links to said assessment form; deliver recurring reminder emails; send emails of encouragement from academic staff or instructors; persuade students that their feedback will be employed both at the institution and course level; offer prizes and other rewards that make survey responses

worth a student's time; teach students how to offer constructive criticism (via their feedback being utilized); and design SET forms that pursue constructive criticism.

Second, Quinn (2002) has quantified eight schemes employed by researchers and institutions that have realized high digitally distributed SET rates of response. Many duplicated or overlapped those previously detailed above, and as a result said researchers unique remaining items are as follows: extend the duration of a given survey's availability; involve students in the choice of optional questions, which makes the SET fundamentally more stimulating to students (this also speaks to Zúñiga's feedback being utilized item); make it clear that all responses are anonymous; make SET forms in a similar fashion to students online assignments; and keep questions and the overall survey brief. Ballantyne (2005) suggested that the effect of both above-mentioned approaches and practices are additive. That is, those individuals or institutions that employ more of these methods will increase their overall response rates of online SET.

Reliability and potential for bias. Currently, there is no clear consensus on the definition of reliability or bias in the student rating domain (Marsh, 1984, 1987; Marsh & Dunkin, 1992). Although, Marsh (1987) states that student ratings are dependable, reasonably effective, multifaceted, relatively unpolluted by potential sources of bias and seen as beneficial by students, faculty, and academic management. His findings also prove that some student scores may have some likely sources of prejudice, unreliability in their opinions, and Edward Thorndike's halo effect. Additional studies agree that SET may be reliable, from the vantage point that students who participate often agree with one another's SET scores (Braskamp & Ory, 1994; Centra, 1993; Marsh, 2007; Marsh & Dunkin, 1992). However, none of the previous texts presume that instructors can be equally effective with students of different ethnic or economic

backgrounds, skills, personality, maturity, and styles of learning. Huff (1954) conceded that SET and extremely consistent ratings most likely do not measure teaching effectiveness; but rather what students fill out on their SET forms, and administrators pretend these items are one and the same. Institutions calculate the resulting statistics, report findings during yearly evaluation cycles, and do not investigate the findings any further.

Instrumentation. Many tools being used for SET lack reliability and valid data (Emerson & Records, 2007). Administration, Faculty, and students believe these to be acceptable, and therefore, they have perceived validity amongst the user base. Ackerman, Gross, and Vigneron (2009) also support that SET have perceived validity via the process of observing multiple instances of instructor teaching throughout a semester, which culminates in a SET being administered.

Normally, SET data is gathered using an assessment form, which contain a fixed number of declarations concerning instructor competencies. Students rate faculty via a Likert-type scale that delivers a systematic measurement range with progressions from strongly disagree to strongly agree. Still, some SET instruments do not follow this stereotypical scoring design. As such, Scriven (1988) created a SET tool that allows students to choose subject prompts, thereby enabling them to note a feature they feel is particularly significant. Not only are the cues divided into two distinct sections; but they also help to outline areas in which faculty are deficient. Scriven (1998) further contends such negative cues aids instructors in improving their teaching practice, while concurrently eliminating the ceiling problem that habitually occurs with Likert-type measures. Students can choose an unlimited number of appropriate cues, and the results are

tabulated by counting the number of times each cue was selected. Altogether, indicating whether a cue is a faculty member's strength or weakness.

SET instruments employed to measure teacher performance also have an impact on the conclusions that can be drawn from the data. Questions must be formulated and dispersed without prejudice, vagueness or inappropriate guidance acting as confounding variables. The ideologies of survey creation have been well defined in multiple studies, and their authors have given great thought to SET questionnaire composition (Aiken, 1996, 1997; Berk, 1979; Oppenheim, 1992). A questionnaires academic legitimacy and validity are determined by substance and instrument (Oppenheim, 1992). Tagomori and Bishop (1995) dissected 200 SET instruments employed in just over 400 colleges of education and uncovered that nearly 58% of the tools contained faults. These included ambiguous items or unclear instructions, response patterns that were skewed or indistinct, and a lack of connection between a given survey question and item that an instructor has purview over or what students might be legitimately expected to assess. Overall, approximately 80% of these SET instruments were comprised of one or more said imperfections. Most tools were custom made and pieced together from unrelated sources, which further displays the need for scrutiny prior to institutional deployment.

Instrument development. To develop a thorough and valid survey instrument, one should employ the following steps: compile and review a great deal of research such as accessible instruments, literature reviews, and correspondence with faculty and students; conduct trials of sample instruments and collect student feedback; and carefully consider the of the psychometric characteristics of the questionnaires during the revision process (Marsh & Hocevar, 1991). Berk

(1979) further elaborates that one must also specify the domain of use prior to SET instrument development as well.

Custom SET instruments. As previously stated, many SET instruments are custom created by an individual institution and as a result, there has not been a wide-ranging and coordinated examination of the reliability and validity of SET. Although many administrators share concerns regarding the legitimacy of SET scores, they are also inclined to emphasize the cumulative rating SET item as the most important measurement of a faculty evaluation (Barth, 2008). Some believe that SET may place too much burden on faculty, rather than students, for student learning and potential for the inflation of grades too (Beran & Violato, 2006).

Not only can an individual quality SET instrument be subject to the variations cited in the previously mentioned paragraphs (faculty, student, and course characteristics), but it can also lend invaluable insight into faculty performance as well. Nevertheless, comprehending how these differences impact SET results are essential to interpreting the data. When utilized appropriately, SET can be a valuable tool for faculty development and administrators to compare faculty, debate for or against tenure and pay increases or promotion (Arreola, 2007).

New instruments. In response to consistent disagreement concerning the reliability and validity of SET, some have tried to build more current and higher quality rating scales. New SET instruments were fashioned in four main studies, and they were reported to be both valid and reliable. They are the Classroom Instruction Evaluation Scale (Emerson & Records, 2007), Evaluation of Teaching Competency (ETCS) scale (Catano & Harvey, 2011), and a Teaching and Learning Quality (TALQ) scale (Frick, Chadha, Watson, & Zlatkovska, 2010). Chiang (2005) altered an existing instrument by developing a scale that distinguishes between effective

and ineffective faculty. In direct comparison to the previously mentioned direct to student tools, another study generated an instrument that measured the effectiveness of student evaluations of faculty to students (Beran, Violato, Kline, & Frideres, 2009)

Each of the measures mentioned above-recognized factors, areas, or classifications pertinent to faculty and their course characteristics. To further elucidate, Chiang's (2005) instrument development method included four classifications: interpersonal relationships, personality characteristics, professional competence, and teaching ability while Emerson and Record's (2007) five attributes were comprised of advocacy, communication, expertise, pedagogy, and professionalism. Also and with student input, numerous factors were applied to teaching effectiveness such as availability, conscientiousness, communication, creativity, feedback, individual consideration, social awareness, problem solving and professionalism (Catano & Harvey, 2011).

Participation by both faculty and students is crucial for SET development and successful implementation. As such, three instruments were created in conjunction with faculty and stakeholder interviews alongside the review of applicable literature. Each tool prevents students from giving a neutral answer via the implementation of an even-numbered Likert scale.

New method of utilization. Rather than endlessly debating whether SET are reliable, valid, or include the proper characteristics, colleges should be discussing the most productive manner in which to integrate SET into their academic systems – to advance both faculty teaching and student learning. Appling, Naumann, and Berk (2001) recommended Triangulation, which would promote a more diverse and consistent evaluation of faculty teaching effectiveness – from the viewpoint of each direct stakeholder: faculty, students, and the individual contributing to the

evaluation. SET are also a valuable component within a three-part system of faculty evaluation. The remaining two are as follows:

1. Observed and mentored by supervisors: As previously mentioned, students are not experts in instructional content or teaching effectiveness experts. Therefore, faculty should also be appraised by peers in their discipline or observed and mentored by supervisors (Appling, Naumann, & Berk, 2001).
2. Self-evaluation of teaching: A self-created portfolio can also help the instructor to demonstrate their capabilities in course content and teaching methods by exhibiting syllabi, defining their educational philosophy, and methods for student engagement.

In all, employing a triangulated system can help to compensate for the limitations of other evaluation types and, when used in conjunction with one another, it can be seen as a good gauge of one's teaching effectiveness.

Peer, administrative, and self-evaluation. Most researchers agree that employing faculty self-evaluations have not been found to correlate highly with peer, administrative, or student evaluations of teaching (Schoofs, 1997). However, faculty peer, administrative and self-assessments seem to largely agree with one another (Centra, 1993). The above-mentioned researchers further state that faculty peer valuations of teaching which include a faculty colleague reviewing a given instructor’s knowledge of the field, quality of course materials, level of rigor, and contributions to program are similar to administrative assessments of teaching and that peer evaluations are highly associated with SET. Thus, the utilization of peer, administrative, and self-evaluations of teaching have not been found to be more useful than

student evaluations, and often, these items were deemed to be less valuable or useful in academic contexts (Seldin, 1984; Centra, 1993).

Conclusion

Teaching is both an art and science with many variables. Thus, it is challenging to design randomized and quality controlled research studies alongside definitive survey instruments. The SET and its processes are synonymous with questions of validity and reliability, and although faculty are aware that characteristics about courses, level of instruction, and student behaviors and traits may impact SET, many faculty are uninformed about what these specific items entail. As many institutions do not regularly observe faculty, students are possibly the lone witness to an instructor's instructional approach, quality of content and grading procedures. Consequently, and despite the perceived number of weaknesses, SET have become a major (if not the only) rubric utilized to inform administrators and subsequently lead to decisions about tenure and promotion. The emphasis on SET to the exclusion of other overarching evaluative measures may result in lost opportunities to improve student learning, overarching faculty irritation and worse yet – high-ranking and talented faculty may resign their posts, which could adversely affect student learning.

Scholarship takes place in a flexible and circumstantial atmosphere. SET can be essential gauges that may aid in students learning - if said pupils are provided with the right questions (Richardson, 2005). As such, this literature review provides the framework for understanding SET, and the possible sources of student and faculty biases related thereto as well.

Chapter III: Methods

Overview

This quantitative study examined what relationship, if any, exists between undergraduate college students' predicted grade and final grade, controlling for the degree of student perception of faculty expertise and perceived affect toward faculty. Also examined is what relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET. These variables measure student biases and perceptions alongside personal sentiment related to faculty expertise and the degree of student learning in both the affective and cognitive domains. Data regarding the degree of student perception of faculty expertise and perceived affect toward faculty were collected cross-sectionally at two points in time, while students' predicted grade was collected at the beginning of the Spring 2018 quarter and actual grade at its end. Data regarding perceptions of faculty knowledge and affect toward faculty (both control variables) are at the level of interval measurement while predicted grade (a predictor variable) and actual earned grade (the outcome variable) in the course are categorical variables. Also, this study examined what relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET. This score is an affective variable representing students' perceptions of their faculty's overall quality and was treated at the interval level of measurement for analysis.

Research Design and Rationale

This study utilized a quantitative, relational, quasi-experimental and observational research design to identify what relationship, if any, exists between the grade undergraduate

college students predict to earn in that course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty.

Faculty knowledge refers instructors' proficiency at constructing circumstances favorable to student comprehension involving both tacit and explicit knowledge together, which is considered a distinctive form of knowledge (Feldman, 1987; Nonaka & Takeuchi 1996). Students predicting their course grade was accomplished by simply asking what letter grade they expect to receive in a given course, a standard approach for measurement of this variable (e.g., Nowell & Alston, 2007). Final course grades (i.e. earned grades) refer to the cumulative results of student performance on assignments and exams after 11 weeks of instruction, letter grades which the instructor submits to the office of the registrar (Harnish & Bridges, 2015).

Perception of faculty knowledge was measured using the Students' Perceptions of Faculty Knowledge (SPFK) instrument (Shih & Chuang, 2013), while students' predicted letter grade was collected via a single question survey: "What Grade do you expect to earn in this course?". Measuring for students' affect toward faculty and course content was accomplished through the administration of The Affective Learning Scale-Abbreviated (ALS-Abbreviated) (Mansson, 2014).

The control variables measure personal biases -- perceptions related to faculty expertise and affect in the affective domain -- while the outcome variable measures the degree of student learning in the cognitive domain. The control variables "perceptions of faculty knowledge" and "affect toward faculty" are at the level of interval measurement, while the predictor variable "predicted grade" and outcome variable "actual earned grade" are categorical variables. Students' predicted grade was solicited in-class during the first day of the term.

Students missing the first day of the course were permitted to complete this measure upon their first class of attendance. Students were also given a two-week period near the end of the term to complete the SPFK and ALS-Abbreviated. Each of the two data collections were administered via Survey Monkey. Email invitations were sent April 23, 2018, with reminder emails sent on April 30, 2018, and May 7, 2018. Students actual earned grade via MI's Office of the Registrar was collected on the last day of classes, May 11, 2018. The overall evaluation they provided their instructor on their end of course SET was also collected on the last day of the course as well.

A quantitative study was chosen to identify what relationships, if any, exist between the grade undergraduate college students predict to earn in a course and their actual earned grade, controlling for students' affect toward faculty and perceptions of faculty knowledge. The relationship between overall SET evaluation and students' earned grade was also analyzed. Such a research design approach for this study was applicable as it explores how the relationships mentioned above may be attenuated by perceived faculty affect and knowledge. Further, and according to Gall, Gall, and Borg (2003), research-based learning on learning outcomes is "an in-depth study of instances of a phenomenon in its natural context and from the perspective of the participants involved in the phenomenon" (p.436). The timeframe selected represented the bulk of an undergraduates Spring term, which is usually when class enrollment is stable, and as a result, response rates are at their highest levels (Akerman, 2009).

Measures. The Students' Perceptions of Faculty Knowledge instrument (SPFK) was created by Shih & Chuang (2013) to evaluate college students' perceptions of faculty knowledge in technology-supported classroom settings. The SPMK is a 49-item questionnaire that utilizes

the following prompts: "Never," "Seldom," "Sometimes," "Often," and "Always," conforming to a 1–5 Likert-type scale. Within the SPFK, faculty are rated within four concepts such as: subject matter knowledge (SMK; "My instructor understands the topics at a high-level ") technological knowledge (TK; My instructor can easily employ technology"), knowledge of students' understanding (KSU; "My instructor is familiar with my prior knowledge in this subject area"), and technological pedagogical content knowledge (TPACK; "My instructor utilizes audio and visual aids and to present intellectual ideas"). The SPFK has a cumulative scoring range from 49 to 245 points, and Table 2 reflects each of the four concepts mentioned above and score ranges as follows:

Table 2
SPFK Conceptual Subscales and Scoring Ranges

Content Area	Number of Questions	Minimum Points	Maximum Points
Subject Matter Knowledge (SMK)	9	9	45
Technical Knowledge (TK)	11	11	55
Knowledge of Students' Understanding (KSU)	6	6	30
Technological Pedagocial Content Knowledge (TPACK	23	23	115
Totals	49	49	245

The goodness-of-fit indices, the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) of the confirmatory factor analysis were used by Shih & Chuang (2013) to examine the construct validity of the SPFK. The RMSEA and SRMR for the underlying four-factor model were 0.089 and 0.083, which indicated a mediocre and acceptable fit of the model to the data, respectively. The person separation

reliabilities of SMK, TK, KSU, and TPACK were computed as 0.95, 0.90, 0.90, and 0.95, respectively (Shih & Chuang, 2013). In a previous sample and study of Taiwanese undergraduate students, all four subscales were also found to be highly reliable (Shih & Chuang, 2013).

For the sake of this research study and resulting data collection population, the Students' Perceptions of Faculty Knowledge instrument (SPFK) was limited to 15 questions, 9 inquiries within the Subject Matter Knowledge (SMK; "My instructor understands the topics at a high-level ") area and 6 inquiries from the knowledge of students' understanding header (KSU; "My instructor is familiar with my prior knowledge in this subject area"). This limited version of the SPMK has a cumulative scoring range from 15 to 75 points, and Table 3 reflects each of the two above-mentioned concepts and score ranges.

Table 3
Limited SPFK Conceptual Subscales and Scoring Ranges

Content Area	Number of Questions	Minimum Points	Maximum Points
Subject Matter Knowledge (SMK)	9	9	45
Knowledge of Students' Understanding (KSU)	6	6	30
Totals	15	15	75

The researcher contacted the authors multiple times for information regarding how the SPMK scores are intended to be interpreted but has hitherto received no reply. As this information is not available, this study adhered to the following interpretation: a score +/1 one standard deviation from the cohorts mean were considered “average,” with lower than one standard deviation considered “low” and higher than one standard deviation considered “high.”

Students' predicted grade was collected via a document that contained a detailed outline of the course grade percentage standards (see Table 4) alongside a single question survey: "What Grade do you expect to earn in this course?"

Table 4

Single Question Survey Grade Scale

Anticipated Letter Grade	Anticipated Percentage Grade
A	90-100
B	80-89
C	70-79
D	60-69
F	0-59

Measuring for students' affect toward faculty was accomplished through the administering of The Affective Learning Scale-Abbreviated (Mansson, 2014), which was developed to measure respondents' affect toward course content and the instructor. Mansson (2014) condensed this form of the original Affective Learning Scale, and it is comprised of 16 items that gauge affect (i.e., positive feeling, attitude, beliefs, and appraisals), eight items regarding class content, and eight items concerning the instructor. Only the eight items related to affect toward the instructor were used in this study, and each reply was measured using a semantic differential scale with a 1-7 point system. Response scores were averaged for each dimension, reported along with standard deviations, then plotted graphically to provide an overarching profile of the connotation of the target concept. Johnson (2009) described reliability coefficients of .82 for affect regarding the course content and .84 for affect regarding the instructor. The coefficient reliabilities were .91 for affect regarding the course content and .94 for affect regarding the course instructor (Mansson, 2014). Per Mansson (2014) and McCroskey (1994), face validity of the tool is admirable. Importantly, the predictive validity thereof is also

convincing. Many studies have utilized this tool and as a result, have produced outcomes that are consistent with the speculative relationships of communicative behaviors with affective outcomes (McCroskey, 1994).

Extant data regarding students actual earned grade, and overall SET evaluation were collected on the last day of classes, May 11, 2018, via MI's Office of the Registrar. The frequencies of both mentioned above and students predicted grades are reported in the study's descriptive statistics.

Population, sampling method, sample, and response rate. The population for this study consisted of undergraduate students enrolled at Musicians Institute, College of Contemporary Music, located in Hollywood, California. As this institution is based on a quarterly academic system and allows for incoming quarterly student (i.e., new) enrollment alongside year-round, full-time study, undergraduate students were considered those who had completed a maximum of 135 total quarterly units and that were enrolled in a minimum of six quarterly units during the Spring 2018 quarter on Musicians Institute's Hollywood campus. First, second, third or fourth-year degree-seeking students may have started their studies in any given quarter, such as Fall, Winter, Spring or Summer – so long as they met the credit requirements above. It is also important to note that many degree-seeking undergraduates enroll in extracurricular performance non-credit courses to gain more vocational industry-relevant performance experience as well. The anticipated age range of the population was 18-24 years, with a high school diploma, GED or international equivalent. The demographics of this population varied regarding socioeconomic status, marital status, dependency, religion, and country of origin.

According to the most recent MI Office of the Registrar enrollment data, the total student population at Musicians Institute was 1,278 for the 2017-2018 year. Of those 1,278 students, 696 students (54.5%) qualified as degree-seeking undergraduate students who completed a maximum of 135 total quarterly units and that were currently enrolled in a minimum of six quarterly units during the Spring 2018 quarter on Musicians Institute's Hollywood campus. The sampling frame was determined via a registrar-generated list of those students who had a minimum of six units in progress during the Spring 2018 quarter and less than 135 completed units. The unit range targeted undergraduate students who are currently enrolled in full-time studies.

For this study, the researcher utilized non-probabilistic sampling. Specifically, to evenhandedly target undergraduate students, the researcher employed total population sampling. This sampling frame provided the researcher with a list of eligible participants to contact via MI's learning management system (LMS) and email to request voluntary participation in the study. To discern a medium size effect of .15 with five predictor variables and thus 73 numerator degrees of freedom, given an alpha of 0.05 and power of 0.80, a minimum of 125 randomly distributed responses were necessary for the planned inferential analyses.

As the LMS invitations and reminder messages did not obtain half of the necessary target sample size by April 6, 2018 (the midway point of data collection), the researcher engaged in increased communication efforts. Increased communication efforts were accomplished through multiple LMS messages and follow-up emails. Had the aforementioned increased communication efforts not resulted in obtaining sufficient participation from the 696 students, the study would have utilized the data received exclusively from the respondents (i.e.,

participants), recognizing that this would have impacted the power of the associated dependent-sample inferential test.

Human subject protections. Written approval for this study was obtained from the Office of Academic Research at Musicians Institute before submitting for approval from the Graduate and Professional School's Institutional Review Board (IRB) at Pepperdine University. Proper licensing for all instruments employed in the study was also obtained via written approval. Participants were provided a letter of informed consent form at the outset of the study, which communicated that participation was voluntary, no compensation for participation would be provided and that they could have withdrawn from participation at any time without consequence.

Digitally signed consent forms were maintained separately from the study's resulting collected data, which itself was retained and kept in secure (encrypted) file. Participant confidentiality was also maintained throughout the process by collecting results via student identification numbers and subsequently assigning a non-identifying number (i.e., code) to each participant for data analysis. This confidentiality was accomplished by generating SHA-256 cryptographic hashes of each student's ID number, yielding a collision resistant (i.e., unique) value which cannot be reverse-engineered to discover the original ID number. Data was reported in aggregate for the study. All data was collected in digital format and will be erased at the conclusion of the above-mentioned three-year period.

Both the risks and benefits of participation were also communicated via the informed consent form before study participation. Risks could have included psychological discomfort and triggers from students reflecting on complex topics such as self-beliefs and faculty perception.

Benefits may have included access to aggregate study results to increase intrapersonal knowledge and non-cognitive academic awareness, as well as a deeper understanding of courses learning objectives and outcomes. Also, the study's findings may be of future use to Musicians Institute's curricular and institutional planning efforts.

FERPA protections. This study followed the Family Educational Rights and Privacy Act (FERPA) prevailing standard for de-identification, thereby following the mantra on whether a "reasonable person in the school community who does not have personal knowledge of the relevant circumstances" could recognize singular pupils based on reasonably accessible data, such as public reports distributed by a governmental organization, or the presentation of meticulous data compiled in tables will a small number of cells (U.S. Department of Education, 2012).

With the above-mentioned items in mind, and in addition to the previously mentioned SHA-256 cryptographic hashes employed to protect each student's ID number, this research study further utilized the U.S. Department of Education's (2012) recommended avoidance of disclosure technique known as blurring. As a result, the reporting of final grades occurred via finite letter grades, which excluded both plus and minus grades and the reporting of exact percentages. Please see table 5 on the following page.

Table 5
Reporting in Letter Grade Format

Reported Letter Grade	Extracted From Percentage Grade
A	90-100
B	80-89
C	70-79
D	60-69
F	0-59

Unblurred reports via cell counts or row and column totals were not published (or made available elsewhere). Digital files will be erased after three years, and any the various copies of said data prepared for the researcher's use were also destroyed immediately after the data analysis phase of this study.

Data collection setting and procedures

Data collection occurred at Musicians Institute, College of Contemporary Music (MI), located in Hollywood, California. Founded in 1977, MI offers Professional Non-Credit, Certificate, Associate of Arts, Bachelors, and Master's Degrees in Music Performance, Composition for Media, and other music industry supportive studies to a student population ranging from 18-24 years, who have previously earned a high school diploma, GED or international equivalent. According to the most recent MI Office of the Registrar enrollment data, the total student population at Musicians Institute was 1,278 for the 2017-2018 year. Of those 1,278 students, data collection was attempted solely from the colleges 696 eligible full-time undergraduate students, those who completed a maximum of 135 quarterly units and that were enrolled in a minimum of 6 quarterly units during the Spring 2018 quarter on MI's Hollywood campus.

Written approval for said data collection was obtained concurrently from MI's Office of Academic Research at Musicians Institute, Board of Directors, and Office of the President. The participants were determined by a registrar-generated list of currently enrolled students who completed a maximum of 135 total quarterly units and that were enrolled in a minimum of 6 quarterly units during the Spring 2018 quarter on Musicians Institute's Hollywood campus. Immediately upon receiving IRB approval, the researcher employed MI's internal Learning Management System (LMS) to request voluntary participation in the study. That is, during each potential participants first quarterly LMS login, a pop-up window appeared to solicit participation, and for those that did agree to partake in said study, they were subsequently and immediately provided with a digital letter of informed consent. This letter was dispersed to both their LMS and individual student email accounts. This electronic document communicated that participation was voluntary, no compensation for participation was provided and that they could withdraw at any time without consequence. After that, each participant was directed via a unique web link where they were instructed to enter their expected grade for the course for which they were enrolled.

Two weeks before the conclusion of the term, participants were also emailed a link to an online survey soliciting their ID number and completion of the Students' Perceptions of Faculty Knowledge (SPFK) instrument and the Affective Learning Scale-Abbreviated (Mansson, 2014). Detailed instructions for successful completion of each instrument was provided, including explicit directions related to the shifting semantic differential scale employed within the Affective Learning Scale--Abbreviated. These emailed invitations were sent on April 23, 2018, with reminder emails subsequently sent on April 30, 2018, and May 7, 2018.

When the collection window expired, the researcher closed the given survey and employed Survey Monkey's reporting features to retrieve the data. Students actual earned grade was collected on the last day of classes, May 11, 2018, via MI's Office of the Registrar.

All electronic data was secured on the researchers' laptop, and backup hard drives via encryption and password protection. During the data analysis phase of the study, access to the files was limited to the researcher alone. All identifying information, including student identification numbers, anticipated grade, and resulting FERPA compliant electronic data will be maintained and properly secured for three years. After that, the data will be erased, and the hard drive(s) will be reformatted.

Analytic techniques. The researcher hypothesized that a positive linear relationship exists between undergraduate college students' perceptions of faculty knowledge, the grade they predict to earn in that course, and their actual earned grade, controlling for students' affect toward faculty. Perception of faculty knowledge, a predictor variable, was measured using the Students' Perceptions of Faculty Knowledge Instrument (SPFK) (Shih & Chuang, 2013), which yields interval level data. Students' predicted grade, a categorical predictor variable, was collected via a single question survey: "What Grade do you expect to earn in this course?". Students' actual earned grade, the categorical outcome variable, was collected via MI's Office of the Registrar. Students' affect toward faculty, the control variable, was measured via The Affective Learning Scale-Abbreviated (Mansson, 2014), yielding data at the interval level of measurement. Table 6 (on the following page) provides a summary of the hypothesis and constituent variables.

Table 6
Hypothesis and Constituent Variables

Hypothesis	Variable	Variable Type	Measure Name	Level of Measurement
1. The relationship between undergraduate college students' predicted grade and final grade is positively attenuated by the degree of students' perception of faculty expertise and their perceived affect toward faculty	1a. Perception of faculty knowledge 1b. The grade students' predict to earn in a given course 1c. Students' actual earned grade in a given course 1d. Students' affect toward faculty	1a. Control 1b. Predictor 1c. Outcome 1d. Control	1a. Students' Perceptions of Faculty Knowledge Instrument (SPFK) (Shih & Chuang, 2013) ** 15 Question Limited Version 1b. Single question survey: " What Grade do you expect to earn in this course?" 1c. Musicians Institute Office of the Registrar 1d. The Affective Learning Scale--Abbreviated (Mansson, 2014).	1a. Interval 1b. Categorical 1c. Categorical 1d. Interval
2. A relationship exists between college students' earned grade and the overall evaluation they provide their instructor on and end of course SET	2a. Students' actual earned g 2b. Overall SET evaluation	2a. Predictor 2a. Outcome	2a. Musicians Institute Office of the Registrar 2b. Musicians Institute Office of the Registrar	2a. Categorical 2b. Interval

Summative data from the single question survey (i.e., "What grade do you expect to earn in this course), and actual earned grade was added manually by the researcher to the Microsoft Excel spreadsheet downloaded from the survey hosting site.

Once the entire data set was manually entered into the above-mentioned Excel spreadsheet, the researcher further employed a secondary assessor to corroborate the data (i.e., to verify that the various figures were entered correctly and that they also match the test results).

Descriptive statistics analysis, multiple logistic regression, and an ANOVA were conducted to determine the objectives of the study. SPSS was used to run the various statistical analyses. A multiple logistic regression was conducted to determine whether a relationship exists between the grade undergraduate college students' predict to earn in a given course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. A level of significance of 0.05 was used.

An ANOVA was conducted to determine whether or not a relationship exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET. A level of significance of 0.05 was used.

Chapter IV: Results

Overview

The purpose of this quantitative, quasi-experimental study was to identify what relationships, if any, exists between the grade undergraduate college students predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. In addition, this study assessed what relationship, if any, exists between undergraduate college students earned grade and the overall evaluation they provide their instructor on an end of course SET. Descriptive statistics analysis, multiple logistic regression, and an ANOVA were conducted to determine the objectives of the study. SPSS was used to run the various statistical analyses. As such, the following research questions and hypotheses were tested in said quantitative analysis:

1. What relationship, if any, exists between the grade undergraduate college students' predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty?

Null Hypothesis. The relationship between college students' predicted grade and actual earned grade is either negatively attenuated by students' perceptions of faculty expertise and their affect toward faculty, or not attenuated at all.

Alternative Hypothesis. It is hypothesized that the relationship between college students' predicted grade and actual earned grade is positively attenuated by students' perceptions of faculty expertise and their affect toward faculty.

2. What relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET?

Null Hypothesis. No relationship exists between college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

Alternative Hypothesis. A relationship exists between college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

Response Rate

Per the most recent MI Office of the Registrar enrollment data, the total student population at Musicians Institute College of Contemporary Music (MI) was 1,278 for the 2017-2018 year. Of those 1,278 students, approximately 54.5%, or 696 students, qualified as degree-seeking undergraduate students who completed a maximum of 135 total quarterly units and were enrolled in a minimum of 6 quarterly units during the Spring 2018 quarter on Musicians Institute's Hollywood campus.

MI's Office of the Registrar provided 696 LMS accounts and e-mail addresses of potential participants. Four of these accounts were undeliverable due to various registrar student account holds, and ten more were hindered by student leaves of absence (LOA), for a total of 682 deliverable student LMS and email accounts (Table 7). 501 students chose to participate in the study for an LMS opt-in rate of 73%. Of those, 441 participants clicked on the link to Shih & Chuang's (2013) Students' Perceptions of Faculty Knowledge instrument and Mansson's (2014) Affective Learning Scale--Abbreviated surveys, for a survey-link click rate of 65%. 398 students began the surveys, with 384 completing at least one of the two surveys. Both surveys were

completed by 346 participants. Those as mentioned earlier yielded a response rate of 69% relative to the number of student LMS opt-ins and an overall response rate of 51%. Two participants who completed the survey did not fit the criteria for participation. After removing these students who did not fit the student's inclusion criteria, a total of 344 sets of usable data remained, for an overall usable response rate of 50%, and a usable response rate relative to the student opt-ins of 68%.

Demographic Information

Eligible student population. The total eligible population for this study consisted of 696 undergraduate college students. The demographics and characteristics of the 696 eligible undergraduate college students are summarized alongside the sample population in Tables 8, 9, and 10. The majority of the 696 eligible undergraduate students were males (532; 76.4%). The mean age among the 696 eligible undergraduate college students was 23.7 years old (SD = 6.58). The oldest of the 696 eligible students was 54 years old, and the youngest was 18 years of age. More than half of the 696 eligible undergraduate college students were non-resident alien (429; 61.6%), with 87 (12.5%) being labeled as Hispanics of any race, and another 104 (15%) identified as White/Caucasian. Half (349; 50.2%) of the 696 eligible undergraduate college students were enrolled in an 18-month Associate of Arts program (AA); 174 (25 %) were enrolled in a 2-quarter Certificate program (C), and 155 (22.2%) were seeking a 12-quarter Bachelor of Music degree. Almost half (313; 45%) of the 696 eligible undergraduate college students were in the first term of their current program, and the highest frequency (216; 31.1%) of 696 eligible students were in their first term at Musicians Institute (i.e., MI Term).

The sample. The final sample of this study consisted of 344 undergraduate college students. The demographics and characteristics of the 344 undergraduate college students are summarized in Tables 8, 9, and 10 alongside the total eligible student population. The majority of the 344 samples were males (264; 76.7%). The mean age among the 344 samples of undergraduate college students was 25.35 years old (SD = 7.02). The oldest of the 344 samples was 54 years old, and the youngest was 18 years of age. More than half of the 344 samples of undergraduate college students were non-resident alien (210; 61%), with 46 (13.4%) being labeled as Hispanics of any race, and another 46 (13.4%) identified as White/Caucasian. Half (172; 50%) of the 344 samples of undergraduate college students were enrolled in an 18-month Associate of Arts program (AA); 87 (25.3%) were enrolled in a 2-quarter Certificate program (C); and 75 (21.8%) were seeking a 12-quarter Bachelor of Music degree. Almost half (151; 43.9%) of the 344 samples of undergraduate college students were in the first term of their current program, and the highest frequency (121; 35.2%) of 344 samples were in their first term at Musicians Institute (i.e. MI Term).

Comparison of populations. Upon reviewing Tables 8, 9, and 10, both the total eligible and final sample undergraduate student populations are similar in key areas. That is, both populations feature a male majority of undergraduate students and the oldest student in each population was 54 years old, and the youngest was 18 years of age. More than half of both undergraduate college populations were non-resident alien, with the next highest frequency being identified as White/Caucasian and the subsequent frequency labeled as Hispanics of any race. In each of the two populations, approximately half of the undergraduate college students were enrolled in an 18-month Associate of Arts program (AA); roughly one-quarter were enrolled in a

2-quarter Certificate program (C), and just over one-fifth were seeking a 12-quarter Bachelor of Music degree. Almost half of both student populations were in the first term of their current program, and the highest number of each of the two student populaces were also in their first term at Musicians Institute. This strengthens the external validity of the study in that the demographics of respondents are quite similar to those of the larger population (which includes students who were eligible to participate in the study but elected not to).

Table 7
Student Opt-in and Response Rates

	Count	% of Students
Total Eligible Students	696	100%
Deliverable Messages	682	98%
Chose to Particpate (LMS Opt-in)	501	73%
Clicked on Survey Link	441	65%
Started Surveys	398	58%
Completed a Section of the Surveys	384	56%
Completed Surveys	346	51%
Useable Survey Data	344	50%

Table 8
Frequency and Percentage Summaries of Gender and Ethnicity

	Eligible Population Frequency	Eligible Population Percent	Sample Population Frequency	Sample Population Percent
Gender				
F	168	24.1	80	23.3
M	532	76.4	264	76.7
Ethnicity				
Asian	19	2.7	12	3.5
Black/African American	25	3.6	11	3.2
Hispanics of any race	87	12.5	46	13.4
Non-Resident alien	429	61.6	210	61
Race/Ethnicity Unknown	42	6	13	3.8
Two or more races	13	1.9	6	1.7
White/Caucasian	104	15	46	13.4

Table 9
Descriptive Statistics Summaries of Age, Cumulative GPA, and Cumulative Comp Credits

	N	Minimum	Maximum	Mean	Std. Deviation
Age of Eligible Population	696	18	54	23.7	6.58
Age of Sample Population	344	18	54	25.35	7.02
Cumulative GPA of Eligible Population	696	0.07	4	3.89	0.76
Cumulative GPA of Sample Population	220	0.09	4	3.44	0.54
Cumulative Comp Credits of Eligible Population	696	0	135	35.12	37.68
Cumulative Comp Credits of Sample Population	344	0	135	36.14	39.22

Table 10
Descriptive Statistics Summaries of Degree, Program Term, and MI Term

	Eligible Population Frequency	Eligible Population Percent	Sample Population Frequency	Sample Population Percent
Degree				
AA - Associate of Arts	349	50.2	172	50
AS - Associate of Science	3	0.4	1	0.3
BM - Bach of Music	155	22.2	75	21.8
C - Certificate	174	25	87	25.3
NC - Non-Credit Extra-Curricular Courses	15	2.2	9	2.6
Program Term				
1	313	45	151	43.9
2	53	7.6	35	10.2
3	111	16	54	15.7
4	61	8.7	33	9.6
5	90	13	49	14.2
6	6	0.8	4	1.2
7	7	1	4	1.2
8	15	2.1	6	1.7
9	6	0.9	2	0.6
11	14	2	5	1.5
12	8	1.1	0	0
13	9	1.3	0	0
14	3	0.5	0	0
15	0	0	1	0.3
MI Term				
1	216	31.1	121	35.2
2	58	8.4	24	7
3	104	15	38	11
4	77	11	51	14.8
5	125	18	57	16.6
6	29	4.1	23	6.7
7	21	3	4	1.2
8	38	5.4	12	3.5
9	6	0.8	2	0.6
13	15	2.1	5	1.5
14	1	0.2	0	0
15	6	0.9	7	2

A majority (298; 86.7%) of the 344 samples of undergraduate college students anticipated their earned grade within a given course to be an A. Nearly half (148; 43%) of

the 344 samples of undergraduate college students earned grades resulted in an A marking. 74 (21.5%) students earned an F grade, while 67 (19.5%) concluded their course with a B grade. Among the 344 samples of undergraduate college students, only 34 (9.9%) or less than 10% had the same anticipated and actual earned grade in their course. The mean cumulative GPA of the 344 student samples was 3.44 (SD = 0.54) with the highest at 4.0 and the lowest resulting in a 0.09. The mean cumulative completed credits of the 344 samples of undergraduate college students was 36.14 (SD = 39.22) with the highest at 185 and the lowest resulting in a 0.0.

Table 11
Descriptive Statistics Summaries of Anticipated and Actual Grades

	Anticipated Grade	Percent	Actual Grade	Percent
A	298	86.7	148	43
B	42	12.2	67	19.5
C	4	0.9	27	7.8
D			28	8.1
F			74	21.5
	No	**Percent**	**Yes**	**Percent**
Equal (Anticipated and Actual Grade)	310	90.1	34	9.9

The scores of the students' perceptions of faculty knowledge and their affect toward faculty are summarized in Table 12 on the following page. Students' perceptions of faculty knowledge was measured using the 15 Question Limited Version of the Students' Perceptions of Faculty Knowledge instrument (SPFK) by Shih and Chuang (2013). The mean score of the 344 samples of undergraduate college students was 53.79 (SD = 13.24). The mean response was in the higher end of the 15 to 75 range of possible scores, which indicated that undergraduate students perceive their faculty to have a great deal of knowledge in the subject for which they

teach. Student's affect toward faculty was measured using the Affective Learning Scale-Abbreviated by Mansson (2014), which was developed to measure respondents' affect toward course content and the instructor. For this study, only the affect toward instructor was used. The mean score was 37.35 (SD = 6.7).

Table 12
Students' Perceptions of Faculty Knowledge and Affect Toward Faculty

	N	Minimum	Maximum	Mean	Std. Deviation
Students' Perceptions of Faculty Knowledge	344	15	75	53.79	13.24
Affect Toward Instructor	344	15	52	37.35	6.7

Results of Hypotheses Testing

Results for research question one. A multiple logistic regression was conducted to determine whether a relationship exists between the grade undergraduate college students' predict to earn in a given course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. The independent variable was the grade students' predict to earn in a given course. The dependent variable was students' actual earned grade in a given course. The control variables were students' perception of faculty knowledge and affect toward faculty. A multiple logistic regression was conducted to test the effect of independent variables on an ordinal measured dependent variable and controlling the effects of covariates. A level of significance of 0.05 was used. The reference category of students' actual earned grade in a given course was F.

The results of the multiple logistic regression are displayed in Table 13. Detailed analysis of the multiple logistic regression had unexpected singularities in the Hessian matrix. This indicates that (1) either some predictor variables should be excluded or (2) some categories

should be merged. As the p-value of the likelihood ratio tests -- using chi-square statistics (χ^2 [28] = 43.44, p = 0.03) -- had a p-value less than the level of significance value of 0.05, the model exhibited a statistically significant fit. Next, the size overall isolated effect of the grade that students predicted to earn in a given course on students' actual earned grade in a given course was estimated.

The Cox & Snell R Square or measure of effect size was 0.54, which reflects a large effect size of the grade students' predict to earn in a given course on students' actual earned grade in a given course -- after controlling the effect of students' perception of faculty knowledge and affect toward faculty. Investigation of the Wald statistic of the individual relationships revealed that none of the various grades students' predict to earn in a given course of A, B, and C had a significant impact on the different students' actual earned grade in a given course of A, B, C, and D -- e.g. when compared to the grade of F, after controlling the effect of students' perception of faculty knowledge and affect toward faculty. As the p-values obtained were all greater than the level of significance value of 0.05, this was an non-significant relationship. As a result, no significant relationship was found between the grade undergraduate college students' predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. Given these results, the null hypothesis for research question one that "The relationship between college students' predicted grade and actual earned grade is either negatively attenuated by students' perceptions of faculty expertise and their affect toward faculty, or not attenuated at all" was not rejected.

Table 13
Multiple Logistic Regression Results

Actual Grade		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
A	Intercept	-24.58	104.21	0.06	1	0.81			
	Students' Perceptions of Faculty Knowledge	-0.01	0.02	0.09	1	0.77	1	0.96	1.03
	Affect Toward Instructor	0.48	0.05	81.58	1	0.00*	1.61	1.45	1.78
B	Intercept	-24.13	142.38	0.03	1	0.87			
	Students' Perceptions of Faculty Knowledge	0.03	0.02	1.94	1	0.16	1.03	0.99	1.08
	Affect Toward Instructor	0.38	0.05	52.64	1	0.00*	1.46	1.32	1.62
C	Intercept	-21.51	241.59	0.01	1	0.93			
	Students' Perceptions of Faculty Knowledge	-0.02	0.02	0.98	1	0.32	0.98	0.94	1.02
	Affect Toward Instructor	0.34	0.06	31.52	1	0.00*	1.41	1.25	1.59
D	Intercept	-12.38	211.35	0	1	0.95			
	Students' Perceptions of Faculty Knowledge	-0.03	0.02	4.46	1	0.04*	0.97	0.94	1
	Affect Toward Instructor	0.07	0.05	2.01	1	0.16	1.07	0.98	1.17

Note:X2(28)= 43.44, p = 0.03, -2 Log likelihood of Reduced Model= 555.71, Cox & Snell R Square = 0.54
a. The reference category is: F.

b. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.

c. This parameter is set to zero because it is redundant.

*Significant at level of significance of 0.05

Results for research question two. An ANOVA was conducted to determine whether or not a relationship exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET. The dependent variables included the overall end of course SET measured using Shih and Chuang's (2013) Students' Perceptions of Faculty Knowledge instrument (SPFK) and their affect toward faculty measured via the Affective Learning Scale-Abbreviated by Mansson (2014). The undergraduate college students' actual grade earned was the independent variable. An ANOVA was employed to test relationships of a continuous measured dependent variable with an independent variable with more than two categorical groupings. A level of significance of 0.05 was used. Results of the ANOVA analysis are displayed in Table 15.

The ANOVA testing exposed that the undergraduate college students' actual grade earned was significantly related to students' perceptions of faculty knowledge ($F(4, 339) = 2.86$, $p = 0.02$), and their affect toward instructor ($F(4, 339) = 77.27$, $p < 0.001$). As indicated by the obtained p-values being less than the level of significance value, significant relationships exist between the independent variable (students actual earned grade) and dependent variables examined (i.e., the overall end of course SET and their affect toward faculty). Given said results, the null hypothesis for research question two "No relationship exists between college students' earned grade and the overall evaluation they provide their instructor on an end of course SET" was rejected.

Post-hoc test results. The post-hoc test results contained in Table 16 display the degree of the relationships of undergraduate college students' actual earned grade with the students'

perceptions of faculty knowledge and their affect toward instructor through an investigation of the differences between groups.

Perceived faculty knowledge. The result of Tukey's test revealed that undergraduate college students who earned a B grade also perceived faculty knowledge as being significantly higher ($p = 0.01$) than those students who earned a D grade (by a mean difference of 9.55 on the scale, which ranged from 15 to 75 possible points). Thus, undergraduate students who earn a higher-level grade concurrently rate their faculty member as having high knowledge, versus those students who earn a low final course grade.

Affect toward faculty. Also, Tukey's test revealed that those undergraduate college students who earned an A grade also have a significantly higher rating of affect toward instructor than those students who earned a grade of B ($p = 0.02$), C ($p = 0.04$), D ($p < 0.001$), or F ($p < 0.001$) by mean differences of 2.25, 2.92, 9.72, and 11.19, respectively, on the scale, which ranged from eight to 56 possible points. Undergraduate college students who earned a B grade also have a significantly higher rating of affect toward instructor than those students who earned a grade of both D ($p < 0.001$) and F ($p < 0.001$) by mean differences of 7.47 and 8.94, respectively. Undergraduate college students that earned a C grade have a significantly higher rating of affect toward instructor than those students who earned a grade of D ($p < 0.001$), and F ($p < 0.001$) by mean differences of 6.80 and 8.27, respectively. Altogether, undergraduate college students who earn higher-level grades also have a higher rating of affect toward instructor than those students who concluded their studies with a lower grade earned.

Table 14
Descriptive Statistics Summaries of Scores of Students' Perceptions of Faculty Knowledge and Affect Toward Instructor by Actual Grades Earned

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Students' Perceptions of Faculty Knowledge	A	148	53.71	13.63	1.12	51.5	55.92	0	60
	B	67	57.37	5.32	0.65	56.08	58.67	38	60
	C	27	51.82	19.13	3.68	44.25	59.38	0	60
	D	28	47.82	16.89	3.19	41.27	54.37	0	60
	F	74	53.66	12.69	1.48	50.72	56.6	0	60
	Total	344	53.79	13.24	0.71	52.38	55.19	0	60
Affect Toward Instructor	A	148	41.22	4.54	0.37	40.48	41.95	31	52
	B	67	38.97	4.95	0.61	37.76	40.18	28	50
	C	27	38.3	4.64	0.89	36.46	40.13	28	47
	D	28	31.5	5.32	1.01	29.44	33.56	21	42
	F	74	30.03	5.34	0.62	28.79	31.26	15	42
	Total	344	37.35	6.7	0.36	36.64	38.06	15	52

Table 15
ANOVA Results of Relationship of Scores of Students' Perceptions of Faculty Knowledge and Affect Toward Instructor with Actual Grades Earned

		Sum of Squares	df	Mean Square	F	Sig.
Students' Perceptions of Faculty Knowledge	Between Groups	1965.17	4	491.29	2.86	0.02*
	Within Groups	58170.91	339	171.6		
	Total	60136.08	343			
Affect Toward Instructor	Between Groups	7338.84	4	1834.71	77.27	0.00*
	Within Groups	8049.6	339	23.75		
	Total	15388.44	343			

*Significant at level of significance of 0.05

Table 16
Post-Hoc Results: Tukey's Test of Difference of Scores of Students' Perceptions of Faculty Knowledge and Affect Toward Inst. by Actual Grades Earned

Dependent Variable	(I) Actual Grade	(J) Actual Grade	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Students' Perceptions of Faculty Knowledge	A	B	-3.66	1.93	0.32	-8.95	1.63
		C	1.89	2.74	0.96	-5.62	9.41
		D	5.89	2.7	0.19	-1.52	13.29
		F	0.05	1.87	1	-5.07	5.16
	B	C	5.56	2.99	0.34	-2.63	13.75
		D	9.55*	2.95	0.01	1.47	17.64
		F	3.71	2.21	0.45	-2.35	9.77
	C	D	3.99	3.53	0.79	-5.7	13.68
		F	-1.85	2.95	0.97	-9.93	6.23
	D	F	-5.84	2.91	0.26	-13.81	2.13
Affect Toward Instructor	A	B	2.25*	0.72	0.02	0.28	4.21
		C	2.92*	1.02	0.04	0.12	5.72
		D	9.72*	1	0	6.96	12.47
		F	11.19*	0.69	0	9.29	13.09
	B	C	0.67	1.11	0.97	-2.37	3.72
		D	7.47*	1.1	0	4.46	10.48
		F	8.94*	0.82	0	6.69	11.2
	C	D	6.80*	1.31	0	3.19	10.4
		F	8.27*	1.1	0	5.26	11.27
	D	F	1.47	1.08	0.65	-1.49	4.44

*. The mean difference is significant at the 0.05 level of significance.

Summary

The purpose of this quantitative, quasi-experimental study was to identify what relationships, if any, exists between the grade that undergraduate college students predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. In addition, this study examined what relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET. Descriptive statistics analysis, multiple logistic regression, and ANOVA were conducted to test the hypotheses posed in this study. Results of the multiple logistic regression showed no significant relationship between the grade undergraduate college students' predicted to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. Results of the ANOVA indicated that a significant relationship exists between college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

Chapter V: Discussion

Overview

The majority of SET related studies repeatedly consider matters related to the creation and validity of an assessment tool (Marsh, 1987), as well as the validity (Cohen, 1981) and reliability (Feldman, 1977) of SET scores. Prior literature focused not only on the effectiveness of instruction, but also the possible sources of student biases related thereto as well (Hofman & Kremer, 1980; Abrami & Mizener, 1983; Tollefson, Chen, & Kleinsasser, 1989). However, few research studies have examined SET and their relation to student learning outcomes. Therefore, this chapter includes an examination of the chief results of this research; a discussion of said findings and their relation to existing SET literature; the complications in this study, and suggestions for future research.

Review of the Results

Results for research question one. The results from research question one suggest that the overall isolated impact of the grade students' predict to earn in a given course on students' actual earned grade in a given course was significant, which indicated that there was a significant relationship between the two aforementioned variables. However, after a detailed investigation of the Wald statistic of the individual relationships revealed that none of the various grades students' predicted to earn in a given course of A, B, and C had a significant impact on the different students' actual earned grade in a given course of A, B, C, and D -- e.g. when compared to the grade of F, after controlling the effect of students' perception of faculty knowledge and affect toward faculty. As a result, no significant relationship exists between the grade undergraduate college students' predict to earn in a course and their actual earned grade,

controlling for students' perceptions of faculty knowledge and their affect toward faculty. Thus, the null hypothesis for research question one was not rejected.

Results for research question two. The results from research question two suggests that the undergraduate college students' actual grade earned was significantly related to students' perceptions of faculty knowledge ($F(4, 339) = 2.86$, $p = 0.02$), and their affect toward instructor ($F(4, 339) = 77.27$, $p < 0.001$). As this study employed a level of significance of 0.05, the results rejected the null hypothesis for research question two: "No relationship exists between college students' earned grade and the overall evaluation they provide their instructor on an end of course SET."

Post-hoc test results. The post-hoc test results illustrate the various degrees of the relationships of undergraduate college students' actual earned grade with the students' perceptions of faculty knowledge and their affect toward instructor through an investigation of the differences between groups.

Perceived faculty knowledge. Tukey's test result revealed that undergraduate college students who earned a B grade also perceived faculty knowledge to be significantly higher than those undergraduate college students who earned a D grade. Thus, undergraduate students' who earn a higher-level grade concurrently rate their faculty member as having high knowledge, versus those students who earn a low final course grade.

Affect toward instructor. Also, Tukey's test revealed that those undergraduate college students who earned an A grade also have a significantly higher rating of affect toward instructor than those undergraduate college students who earned a grade of B. Undergraduate college students who earned a B grade also have a significantly higher rating of affect toward instructor

than undergraduate college students who earned a grade of either D or F. Undergraduate college students that earned a C grade have a significantly higher rating of affect toward instructor than those undergraduate college students who earned a grade of D or F. Altogether, undergraduate college students who earn higher-level grades also have a higher rating of affect toward instructor than those undergraduate college students who concluded their studies with a lower grade earned.

Conclusions

Research question one. As previously indicated, this study found no significant relationship exists between the grade undergraduate college students predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward faculty. Although Matos and Ragan (2010) share that much earlier research has studied the correlation between predicted course grade and SET, assessing students' final course performance in relation to students predicted grade is much overlooked in the current literature (Marsh and Roche, 2000). As a result, this study aids in supplementing and possibly extending the suggestions of earlier researchers who advocate for SET assessment systems that simultaneously assess both student and instructor achievement. For example, those learner-centric evaluation processes where students assess their own coursework and subsequently predict their final course grade alongside evaluating the instructional quality of their assigned faculty member might provide a more encompassing view of teaching effectiveness (Cohen, 1990; Kwan, 1999; Whitworth, Price, & Randall, 2002).

The majority of studies related to students predicted grade in higher education have also been regulated to either (1) a given researcher seeking to predict students' grades in relation to

various faculty behavioral patterns, select course attributes, or classroom environments; or (2) the researcher employing machine learning alongside historical grade data to predict students' course grades for a future enrollment term. Thus, said foci has steadily neglected research that investigates students predicting their grade in relation to both their final grade or the overall evaluation they provide their instructor on an end of course SET.

Also, discussed in the findings above, the overall isolated impact of the grade students predict to earn in a given course on students' actual earned grade in a given course was significant, which indicated that there was a significant relationship between the two aforementioned variables as well. However, after controlling for (i.e., removing) the effect of students' perception of faculty knowledge and affect toward faculty, the relationship between college students predicted grade and actual earned grade was not attenuated at all. Thus, and although the seminal works of Marsh (1987) and Marsh and Dunkin (1992) suggest no clear consensus on the definition of bias in the student rating domain, this study's select findings and control variables suggest otherwise. Namely, that items such as students' perceptions of faculty knowledge and affect toward faculty are associated with students learning outcomes and final course grades. While Beleche, Fairris, and Marks (2012) share that grades are not perceived as one's full measure of learning, this study is an important step in extending the literature employed to link SETs and student perceptions of faculty to final student grades and course learning outcomes.

Research question two. The results from research question two suggests that the undergraduate college students' actual grade earned was significantly related to those students' perceptions of faculty knowledge and their affect toward instructor. It is well established that

anticipated grades frequently correlate with positive ratings of professors, and that rigor is often negatively related or believed to be negatively related to SET (Clayson, 2009; Clayson & Haley, 2011; Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Sojka, Gupta, & Deeter-Schmelz, 2002). In addition, students who anticipate earning a high grade are more likely to deliver higher ratings (Barth, 2008; Beran & Rokosh, 2009b; Beran & Violato, 2006; Beran & Violato, 2009; Olshavsky & Spreng, 1995; Nowell et al., 2010; Serdyukova, Tatum, & Serdyukov, 2010). However, students actual earned grade within said SET literature is often ignored.

As such, the findings of this study may extend the current SET literature by helping to link student learning outcomes (i.e., grades) to students' overall evaluation they provide their instructor on an end of course SET. Specifically connecting student's (1) perceptions of faculty knowledge, (2) the value of said instructor and (3) and the given likelihood of taking future courses with one's specific faculty member to student learning outcomes. To further elucidate and while keeping research question two's findings in mind, one could surmise that students who perceive their instructor to have high knowledge and value within the classroom may pay greater attention to faculty in-class lectures, do homework to greater quality, engage in meaningful discussions both in-class and with their peers related to the class subject matter, and ultimately learn more in their classes. Altogether, these actions may plausibly result in an elevated end of course grade.

Post-hoc testing reinforcement. The findings mentioned above are further reinforced by the post-hoc test findings as well. Specifically, undergraduate students' who earned a higher-level grade concurrently rate their faculty member as having high knowledge, versus those students who earn a low final course grade. Also, undergraduate college students who earned

higher-level grades also have a higher rating of affect toward instructor than those undergraduate college students who concluded their studies with a lower grade earned.

SET leniency hypothesis. The findings within both research question two and the post-hoc testing also contradict a popular premise found in much of the SET literature reviewed, specifically, the notion that a "leniency hypothesis" or paradox of rigor exists. This theory assumes that instructors who employ a lenient grading scheme may gather more favorable SET and as a result achieve superior overall evaluation scores (Gump, 2007; Heckert, Latier, Ringwald-Burton, & Drazen, 2006). However, this study's findings suggest that instructors who are perceived as knowledgeable and of having value entice students to achieve higher grades, while concurrently granting high SET scores to said faculty.

Additional contributions of the study. There were other contributions to this study, including a high response rate and reliable and valid SET instrument utilization.

Response rate. Of the 682 students eligible to participate in this study, both surveys were completed online by 344 participants, for an overall response rate of 50%. Many prior studies, however, suggest that online SET methods averaged a 33% rate of response compared to the 56% attained in paper dispersals (Richardson, 2005). Other studies suggest that response rates for online evaluations are even lower than for paper-and-pencil in-class evaluations by as much as 30% (Avery, Bryant, Mathios, Kang, & Bell, 2006; Dommeyer et al. 2004; Guder & Malliaris 2013). As students in classes with low enrollment might feel that their secrecy is more questionable, small classes are also known to have radically lower SET response rates than evaluation administered in larger classes (Feldman, 1984; Marsh, 1987; Perry & Smart, 2006).

Therefore, the 50% response rate in this study is particularly salient for two main reasons. First, this study exceeds typical digitally dispersed response rate by nearly 20%. Second, MI classes conventionally enroll no more than 9 to 15 students, and as a result, these small class sizes did not adversely affect this study's response rate. However, it is unclear if these abnormalities are due to the very nature of MI students or that the researcher followed many expert's best practices to increase student response rates. On one hand, it is known that musicians generally tend to be both highly engaged in their studies and very competent in using technology (Schultz, 2013). On the other, the researcher employed many of the US Teaching and Learning with Technology/Flashlight Group's best exercises for increasing digitally dispersed SET response rates as well (Zúñiga, 2004).

Instruments employed. A questionnaire's legitimacy and validity are informed by substance and an instrument's psychometric properties (Oppenheim, 1992). Tagomori and Bishop (1995) dissected 200 SET instruments employed in just over 400 colleges of education and uncovered that nearly 58% of the tools contained faults. These included ambiguous items or unclear instructions, response patterns that were skewed or indistinct, and a lack of connection between a given survey question and item that an instructor has purview over or what students might be legitimately expected to assess. Overall, approximately 80% of these SET instruments were comprised of one or more such imperfections. Most tools were custom made and pieced together from unrelated sources, which further displays the need for scrutiny before institutional deployment.

However, the instruments employed within this study to measure students' perception of their instructor's knowledge and affect are deemed useful to properly evaluating faculty, reliable,

and valid (Johnson, 2009; McCroskey, 1994; Shih & Chuang, 2013; Mansson, 2014). All four subscales utilized in Shih and Chuang's (2013) Students Perceptions of Faculty Knowledge (SPFK) instrument were found to be highly reliable, and the confirmatory factor analysis for the underlying four-factor model indicated a mediocre and acceptable fit of the model to the data, respectively. Per Mansson (2014) and McCroskey (1994), face validity of the Affective Learning Scale - Abbreviated is admirable and predictive validity thereof is convincing. Many studies have utilized this tool and as a result, have produced outcomes that are consistent with the speculative relationships of communicative behaviors with affective outcomes (McCroskey, 1994). As such, the above-mentioned tools could be employed in lieu of more traditional SETs to make critical judgments regarding faculty course assignments, tenure, and possible positional promotion alongside the systemic demonstration of institutional effectiveness to various higher education accrediting agencies.

Also, there is no consensus on (or commonly accepted definition of) what "good" teaching is, nor has an entirely agreeable standard of teaching effectiveness been established (Adams, 1997). As a result of the items above, this study has the potential to positively impact those academic administrative staff who seek to measure students feedback directly related to perceived faculty knowledge and instructional value; as well as require usable SET findings and high response rates to employ time-tested and valid instrumentation.

Additional limitations in the study. In addition to the assumptions, limitations, and delimitations discussed within Chapter I, the preceding findings of this study are based on numerous caveats. First, the resulting figures rely solely on surveys submitted by students who participated in this research. If such students did not mimic a random sample of the students who

enroll in a given class type, the results could be impacted by the selection of this sample population. Naturally, this is an inherent challenge contained within all studies involving voluntary participation. That is, practically all researchers are unable to obtain the information of those students who do not complete SET, which makes it all but impossible avoid sample selection bias lest 100% of students elect to participate in the study.

Additional studies contain the response rate of students as an explanatory variable, which may have the ability to control for the selectivity of sample. Thus, students who do not participate in SET might tend to score the faculty member differently than those undergraduates that participate in the SET process. Alternately, Ragan and Walia (2010) contend the quality of the instructor directly impacts student response rate, and as a result, arguing that it should not be included as a control variable.

Suggestions for Further SET Research

Although much of the current SET literature has focused on student biases towards faculty, faculty biases of the SET administrative review process, impacts of classroom environments, and attributes of a given course, this research study sought to solely evaluate student learning outcomes while controlling for SET items such as students' perceptions of faculty knowledge and their affect toward faculty. As such, this study did not include whether an instructor’s race, physical appearance, or class meeting time impacts perceptions of faculty knowledge, students affect toward instructor and the overall result of a given end of course SET. Furthermore, as this study has not adequately controlled for course level, class size, prior student interest and student age, future SET research could benefit from employing such tactics to expand this studies method as well.

As Ramsden (1991) states, the subject matter presented in a given class also influences student SET scores, researchers within specified music or other disciplines may wish to further investigate whether or not course attributes would be likely to affect a given departments end of course SET scores – i.e. in the way the current studies have projected, in an altered fashion, or if they display no affect at all. Further, the scope of this research did not include General Education courses offered offsite via the college's various articulation agreements with regionally accredited institutions, for which second, third or fourth-year baccalaureate students may be enrolled. As a result, expanding the scope of this study to include said courses may enrich the data set related thereto. Additional research regarding course subject matter and its impact on students' perceptions of faculty knowledge, students affect toward instructor, and the overall result of a given end of course SET would also be welcomed.

Further, and given the venue of this research study, both required and elective courses in lecture and musical performance formats were included in this study. As each course grade type was treated equally, future studies could benefit from accounting for such items. By doing so, this could also help to either reinforce or challenge conventional literature. To further clarify, Feldman (2007) shared that required courses obtain lower student ratings than electives and both Fiske (1977) and Kaiser (1998) asserted that no relationship has been found between an evaluators reliability in adjudicating student musical performances and their performing ability as measured by applied musical grades.

A great deal of research also suggests that higher level classes receive superior SET scores (Feldman, 1978; Marsh, 1987). As a result, studying the impact of course level alongside

elective versus required courses, and lecture versus musical performance (or other participatory type) classes may also expand researchers understanding of SET.

Anxiety of student. Select courses, such as those that required public student musical performances or a good amount of student in-class participation, may have led some individuals to experience higher levels of anxiety than those courses that were lecture and homework based alone. As a result, it may be beneficial to conduct additional studies to determine the impact these feelings may have had on students in this study's (and future studies of) student evaluations of teaching, including students' perceptions of faculty knowledge and their affect toward faculty. Additional research could also be conducted on how said anxieties may impact a given individual's anticipated course grade, weekly assignment, and final course grades as well.

Level of student. First and second-year students are thought to rate their instructors more critically than those in years three and four (Clayson, 2009). Numerous studies also suggest that students who have prior knowledge in the subject area preceding a given course usually provide higher SET scores (Feldman, 1977, p. 236; Marsh & Cooper, 1981; Prave & Baril, 1993). In addition, almost half (151; 43.9%) of the 344 samples of undergraduate college students in this study were in the first term of their current program (i.e., a new subject matter), and the highest frequency (121; 35.2%) of the 344 samples were in their first term at Musicians Institute (i.e., first-year students).

With this in mind, this study did not specifically compare student perceptions of faculty and their affect toward instructor with student level or prior student knowledge. Thus, it would be interesting to see if first-year students or those unfamiliar with their given subject matter (i.e., first MI term students) rated their instructor any differently than those in subsequent terms.

Furthermore, it may also be helpful to account for (or control for) students prior knowledge via an appropriate evaluation instrument (i.e., control variable) that measures students' knowledge or ability.

Student GPA and self-esteem. To review, the majority (298; 86.7%) of the 344 samples of undergraduate college students anticipated their earned grade within a given course to be an A. However, only half (148; 43%) of the 344 samples of undergraduate college students earned grades resulted in an A marking. Also, only 34 (9.9%) students had the same anticipated and actual earned grade in their course. As such, future studies could benefit from including participants grade point average (GPA) at the start of the study and grade that they earned in a previous serial course. To further clarify, this may help to determine if undergraduate students predicted grade was somewhat realistic. That is, a student who receives a D grade in a level one music theory course, who subsequently participates in a study such as this one may not have a high probability of earning an A marking in their level two offering.

Further, students who have a high amount of self-esteem do not usually possess elevated levels of academic performance and said confidence levels may sometimes be counterproductive (Baumeister, Campbell, Krueger, & Vohs, 2003). With the abovementioned high amount of predicted A grade markings in mind, it may also be interesting to determine if a correlation exists between elevated levels of self-esteem, academic ability or actual earned grades in relation to this SET research.

Student sentiment. Another potentially meaningful future research study may be to include methods that measure student sentiments related to the usefulness of SET. Given the

high response rate within the present study, it may be of value to investigate undergraduate's thoughts regarding the utility SET scores as a possible explanation for such a high response rate.

International student population. As previously discussed, MI's population is composed of approximately 40% international students and nearly 62% of the undergraduate students in this study were classified as non-resident alien. Also, many colleges and universities are focusing their enrollment efforts on international students, which aids in supplementing the gap between state funding and students' inability to secure Title IV Financial Aid funding such as the Federal Pell Grant, Federal Supplemental Educational Opportunity Grant (SEOG), Federal Perkins Loan, Federal Subsidized and Unsubsidized Direct Loans (Bound, Braga, Khanna, & Turner, 2016). This trend may be exacerbated by the current immigration practices of the Trump administration which has resulted in a 40% decline in international enrollees nationwide (Zhou, 2018). However, many scholars focus on the recent Department of Education data that suggests foreign-student enrollment has risen consistently for the last 35 years and that this trend will continue into future enrolment cycles (Hussar & Bailey, 2017).

Further, as many cultures outside the United States revere both teachers and elders, comprehending international student assimilation challenges has far reaching implications for SET (Andrade, 2006). Therefore, and with each of the items above in mind, a differential analysis of international students versus domestic students would be welcomed in future studies.

Faculty and administrative staff. To establish how frequently and what types of modifications instructors make to their material based on SET results, further studies of faculty reaction to the results of students' perceptions of faculty knowledge, students affect toward instructor, and the overall result of a given end of course SET may be useful as well. There is

also potential to assess how collegiate administrative staff utilize the results of this study or of Shih and Chuang's (2013) Students' Perceptions of Faculty Knowledge instrument (SPFK) and Mansson's (2014) Affective Learning Scale-Abbreviated tools to make critical judgments regarding faculty course assignments, tenure, and possible positional promotion alongside the systemic demonstration of institutional effectiveness to various higher education accrediting agencies.

Faculty tenure and preparation. Centra (1981) suggested that faculty members who have regularly been teaching for three to 12 years are more likely to receive high SET scores than those instructors who do not hold experience within said ranges. However, notwithstanding the studies that contained teaching assistants, Feldman (1983) found no significant relationship between faculty rank and SET ratings. Thus, future research related to perceptions of faculty knowledge, student affect toward instructor and SET scores could further account for instructor tenure, experience level, and rank.

Also, good evaluations can result from those faculty who are organized, deliver accurate and uncluttered lectures and communicate in a clear manner (Donnon, Delver, & Beran, 2010). Thus, it would be interesting to determine if the amount of time faculty prepare in advance for a given course correlates with students' perceptions of faculty knowledge and affect toward instructor as well.

Faculty teaching style. Also, when students are not presented with opportunities to participate in a given class, instructor self-expression and overall teaching style had a more significant impact on SET scores than lecture content; however, when opportunities were shared, expressiveness was far less significant (Perry, Abrami, Leventhal, & Check, 1979; Marsh &

Ware, 1982). With these items in mind, it may also be of use to include the teaching style of faculty as a covariate within another study that follows the parameters and methods of this research.

Faculty grade distribution. In a grade distribution experiment, Powell (1977) utilized a trio of dissimilar grading rubrics, classified as stringent, moderate, and lenient, in separate divisions of identical classes. Each of the grading benchmarks caused markedly different grade distributions in the expected manner. That is, assessment of the value of faculty instruction diminished as the toughness of the grading standards increased. Langbein (2008) a faculty member at a mid-sized college further shared that a new consumer model of higher education results in routine grade redistribution and inflation, both for favorable SET scores and to avoid time-consuming student complaints after the close of a semester. As such and in order to ensure that each students' actual earned grade is treated equally, it may be beneficial to include each instructors grade distribution schemes as an additional control variable or further account for such attributes within this type of study.

Peer evaluation. The utilization of peer evaluations of teaching which include a faculty colleague reviewing a given instructor's knowledge of the field, quality of course materials, level of rigor, and contributions to program have not been found to be more useful than student evaluations (Centra, 1993). Oftentimes, these peer instructional assessment items were deemed to be less valuable or useful in academic contexts (Seldin, 1984; Centra, 1993). However, it would be interesting to determine the correlation between students' perceptions of faculty knowledge, students affect toward instructor and peer perceptions of faculty knowledge and affect toward instructor.

Institution and timeframe. It was beyond the scope of this study to look at all types of colleges and universities. As such, the researcher focused on student evaluations of teaching at Musicians Institute, a small private undergraduate degree-granting music college, located in Hollywood, California. Thus, it would also be interesting to learn if this study's results would vary contingent upon whether it was conducted at a two-year community college, average four-year university or another type of specialized research or vocational based institution.

As data collection occurred during MI's spring quarter, the timeframe of conducting said research was also another key restriction imposed by the researcher. Thus, this researcher would also like to conduct this study during the college's summer, fall, and winter quarters to determine the impact of seasonality on SET results. After conducting such analyses, future research could attempt to track undergraduate student results from year one to year two; year two to year three; and from year three through graduation at the conclusion of year four. It would also be helpful to study the variance in student age (or associated maturity thereof) at the time the assessments are dispensed, which could be a larger source of influence on SET scores than the timing of the evaluations themselves (Centra, 1993; Feldman, 1978).

Comparison of online and paper responses. Various researchers believe that asynchronous SET questionnaires lead to either harsher or more compassionate assessments, alongside wholly inaccurate or diminished faculty ratings, and lower response rates (Nowell et al., 2010; Serdyukova et al., 2010). Thus, it may also be beneficial to conduct this study again and employ both paper and digital dispersals of Shih and Chuang's (2013) Students' Perceptions of Faculty Knowledge instrument (SPFK) and The Affective Learning Scale-Abbreviated

(Mansson, 2014). By dispersing both paper and digital surveys, future researchers would be able to compare the results and response rates of this study in relation to each survey type as well.

Conceptual framework. SET appears to be devoid of a prevalent theory. Due to the concentration on the correlation between theoretically biasing variables and student course ratings, numerous researchers employ regression methods and design their analyses in comparable manners. These methods are usually accomplished through the constructing of predictive models from the vast amount of empirical data. As such, the creation of a robust conceptual framework via future SET studies could potentially be a significant contribution to the body of SET literature. Such a framework could be used in conjunction with the large amount of SET literature to inform, update, and further future research.

Suggestions for Policy and Practice

Administrative staff. The following policy and practice suggestions are intended for those administrative staff who disperse SET instruments to students and oversee faculty.

Student SET onboarding. The highest frequency of the aforementioned 344 sample learners (121; 35.2%) were in their first term at Musicians Institute (i.e. MI Term). As a result, said undergraduate collegiate students had not previously participated in the evaluation of faculty within a collegiate learning environment. Legitimate peripheral participation provides a method to speak properly about crucial academic matters between students, faculty and administrative staff about their class activities, artifacts, knowledge and practice (Lave & Wenger, 1991). As such, Jonassen and Land (2012) share that today's inbound student body requires guidance in understanding the importance of academic interactions, and their subsequent role in providing value-based feedback to provide measurable metrics and gauge faculty performance within their

new collegiate community. Thus, incoming undergraduate students would benefit from an introductory scaffolding collegiate SET onboarding course. Upon completing such a collegiate administrative initiative (i.e. course), students should be able to demonstrate the following learning objectives:

- Identify best practices within a given professor's classroom management and instructional techniques.
- Comprehend and recognize potential sources of student biases toward faculty and the impacts of a given classroom environment on SET and student learning outcomes.
- Participate in detailed evaluations of a professor's communication skills, organization, enthusiasm, flexibility, knowledge of the subject matter, clarity, course difficulty, and fairness of grading.
- Give professors a numerical rating on each of the aforementioned items.
- Provide measurable and useful feedback that can be used to improve the faculty members level of instruction.

Student focus. Although Marsh and Dunkin (1992) agree that bias does not exist in the domain of SET, this study's findings suggest that items such as students' perceptions of faculty knowledge and affect toward faculty are associated with students learning outcomes and final course grades. Not only would this be helpful for students to receive this information at the outset of a given course or semester; but also, this information would potentially enable each pupil to control for such predispositions in their daily coursework. Thus, said students may be able to exceed their current academic performance and earn elevated grade markings and ultimately learn more within their studies.

In addition, students have also testified that they gave their faculty member a higher or lower evaluation on a semester-ending SET than the instructor deserved (Clayson & Haley, 2011). As such, administrative staff could also benefit from communications that encourage students to communicate on their SET in an honest and forthright manner.

Faculty focus. Numerous faculty members believe that SET scores are not reliable and that such pupil feedback does nothing to benefit improve their instructional practice (Balam & Shannon, 2010; Beran & Rokosh, 2009). Nearly all instructors also believe that students are not specialists in classroom teaching and as a result, they can only judge their reaction to a course (Ackerman, Gross, & Vigneron, 2009). Yet, when faculty are well informed about their given institution's reasons for evaluation and the processes related thereto, much of their uneasiness diffuses and the willingness to learn from undergraduate feedback increases (Sojka, Gupta, & Deeter-Schmetz, 2002; Hativa, 1995; Gallagher, 2000; Bain, 2011).

With the items above and recommendations within the administrative student focus header above in mind, managerial and directorial collegiate staff could benefit from routine detailed SET communications to faculty. That is, both documents and departmental meetings that clearly outline items including (but not necessarily limited to): the institution's willingness to employ valid SET instruments that enable reliable data collection alongside measures that control for potential student biases; the dispersal of a student SET onboarding course that will encourage students to both take the SET process seriously, while concurrently educating said pupils on best practices in classroom instruction; and the SET reminders sent to students each semester that clearly communicate that students' perceptions of faculty knowledge and affect toward faculty are associated with students learning outcomes and final course grades. Not only

should these items have the potential to mitigate faculty concerns; but also, they will arm administrative employees with a more robust data set – i.e., a data set that can potentially help them to make more informed decisions regarding wage increases, course assignments, retention, promotions, and tenure (Seldin, 1993; Thompson & Serra, 2005).

Faculty recommendations. As previously discussed, the findings from research question two suggest that the undergraduate college students' actual grade earned was significantly related to students' perceptions of faculty knowledge and their affect toward instructor. Said findings were further reinforced via the post-hoc test findings as well. Specifically, that undergraduate students' who earned a higher-level grade concurrently rate their faculty member as having high knowledge, versus those students who earn a low final course grade. Also, undergraduate college students who earned higher-level grades also have a higher rating of affect toward instructor than those undergraduate college students who concluded their studies with a lower grade earned.

Wood, Bruner, and Ross (1976) further defined the role of the instructor as expert. As such, it would be helpful for faculty to routinely communicate their domain specific expertise and unique professional experiences in demonstrative terms as a scaffolding of sorts (e.g., within lectures and class meetings), which would have the potential to elucidate students on faculty members breadth of knowledge. Altogether, this may result in higher SET scores, superior levels of students' perceptions of faculty knowledge, and elevated markings of affect toward instructor, thereby resulting in better actual earned grades and subsequently meeting published course student learning outcomes.

Student recommendations. While no reliable association exists amongst student characteristics and SET, many studies agree that students are perfectly situated to give feedback regarding their experience in a course (Abrami & Mizener, 1983). Thus, it is imperative that students take the administrative and faculty communications discussed above seriously and as a result, that they subsequently and fully immerse themselves in all pedagogical matters, SET related onboarding courses and the like.

Recommendations for those outside collegiate programs. The findings in this study may also be of use to those educators outside the collegiate music school domain, which may include private lessons programs, community workshops, afterschool programs and the like. Particularly those items related to the impact of students' perception of faculty knowledge and affect toward instructor on student learning outcomes. As a result, those individuals that oversee such programs may wish to invest time, effort, and resources into educating students on a particular instructors' background and expertise. Not only to entice students to enroll in (or continue in) a given program or lesson series; but also to ensure that each pupil is given the maximum opportunity to achieve their full learning goals.

Per the collegiate faculty recommendations above, it would also be helpful for lesson, workshop, and afterschool instructors to routinely communicate their subject matter expertise, educational background and unique professional experiences in demonstrative terms to students, which again - may result in higher levels of students perceptions of instructor knowledge, and elevated markings of affect toward instructor, thereby resulting in better student learning outcomes. Further, and if students continually meet their desired learning goals, these above-mentioned recommended methods to learning program leaders may have the potential to inspire

students that did not previously enjoy organized learning (i.e., school) to pursue lifelong learning. Not only during their K-12 years but also, potentially pursue collegiate studies as well.

Summary

The majority of student evaluations of teaching (SET) related studies repeatedly consider matters related to the creation and validity of an assessment tool, as well as the validity, and reliability of SET scores. Not only to determine the effectiveness of instruction; but also, the possible sources of student biases related thereto as well (Hofman & Kremer, 1980; Abrami & Mizener, 1983; Tollefson, Chen, & Kleinsasser, 1989). However, few research studies have examined SET and their relation to student learning outcomes. Therefore, the purpose of this study was to identify what relationships, if any, exists between the grade undergraduate college students' predict to earn in a course and their actual earned grade, controlling for students' perceptions of faculty knowledge and their affect toward instructor. Also, this study also examined what relationship, if any, exists between undergraduate college students' earned grade and the overall evaluation they provide their instructor on an end of course SET.

The population for this study consisted of 344 undergraduate college students enrolled during the spring 2018 quarter at a small private college of music, located in Hollywood, California. A detailed investigation of the Wald statistic of the individual relationships revealed that none of the various grades students' predicted to earn in a given had a significant impact on the prediction of the different students' actual earned grade in a given course after controlling the effect of students' perception of faculty knowledge and affect toward faculty. However, the undergraduate college students' actual grade earned was significantly related to students' perceptions of faculty knowledge and affect toward faculty. The findings mentioned above are

further reinforced by the post-hoc test findings too. Specifically, that undergraduate students' who earn a higher-level grade concurrently rate their faculty member as having high knowledge, versus those students who earn a low final course grade. Also, undergraduate college students who earn higher-level grades also have a higher rating of affect toward instructor than those undergraduate college students who concluded their studies with a lower grade earned.

With the above-mentioned items in mind, it is hoped that future scholars will be able to utilize the foundation established within this study to further evaluate the relationship between course, student, students' perceptions of faculty, and student evaluations of teaching. While the findings within this study may not be unanimously applicable to the collegiate population at large, said results can function as additional data to be contemplated in measuring the impact of students' perceptions of faculty knowledge and their relation to learning and grade outcomes.

REFERENCES

Abbott, R. D., Wulff, D. H., Nyquist, J. D., Ropp, V. A., & Hess, C. W. (1990). Satisfaction with processes of collecting student opinions about instruction: The student perspective. *Journal of Educational Psychology*, *82*(2), 201. https://doi.org/10.1037/0022-0663.82.2.201

Abrami, P. C. (1989). How should we use student ratings to evaluate teaching? *Research in Higher Education*, *30*(2), 221-227. https://doi.org/10.1007/BF00992718

Abrami, P. C., & Mizener, D. A. (1983). Does the attitude similarity of college professors and their students produce "bias" in course evaluations?. *American Educational Research Journal*, *20*(1), 123-136. https://doi.org/10.3102/00028312020001123

Abrami, P. C. (1989). Book Reviews: Seeing the Truth About Student Ratings of Instruction. *Educational Researcher*, *18*(1), 43-45. https://doi.org/10.3102/0013189X018001043

Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: what we know and what we do not. *Journal of educational psychology*, *82*(2), 219-231. https://doi.org/10.1037/0022-0663.82.2.219

Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, *72*(1), 107-118. https://doi.org/10.1037/0022-0663.72.1.107

Abrami, P. C., Perry, R. P., & Leventhal, L. (1982). The relationship between student personality characteristics, teacher ratings, and student achievement. *Journal of Educational Psychology*, *74*(1), 111-125. https://doi.org/10.1037/0022-0663.74.1.111

Ackerman, D., Gross, B. L., & Vigneron, F. (2009). Peer observation reports and student evaluations of teaching: Who are the experts? *Alberta Journal of Educational Research*, *55*(1), 18–39. Retrieved from Eric database. (EJ838670)

Adams, J. V. (1997). Student evaluations: The rating game. *Inquiry*, *1*(2), 10-16. Retrieved from Eric database. (EJ591701)

Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behavior, personality, and attitudes*. Oxford, England: John Wiley & Sons.

Aiken, L. R., & Aiken, L. R. (1997). *Questionnaires and inventories: Surveying opinions and assessing personality*. New York, NY: John Wiley.

Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 110-145). Beverly Hills, CA: Sage.

Aleamoni, L. M. (1987). Student rating myths versus research facts. *Journal of Personnel Evaluation in Education*, *1*(1), 111-119. https://doi.org/10.1007/BF00143282

Aleamoni, L. M., & Hexner, P. Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science*, *9*(1), 67-84. https://doi.org/10.1007/BF00118969

Allen, M. (1996). Research Productivity and Positive Teaching Evaluations: Examining the Relationship Using Meta-Analysis. *Journal of the Association for Communication Administration, 2*(1), 77-96. Retrieved from Eric database. (ED379705)

Al-Maamari, F. (2015). Response Rate and Teaching Effectiveness in Institutional Student Evaluation of Teaching: A Multiple Linear Regression Study. *Higher Education Studies*, *5*(6), 9-20. https://doi.org/10.5539/hes.v5n6p9

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of personality and social psychology*, *64*(3), 431-441. https://doi.org/10.1037/0022-3514.64.3.431

Anderson, H. M., Cain, J., & Bird, E. (2005). Online student course evaluations: Review of literature and a pilot study. *American Journal of Pharmaceutical Education*, *69*(1), 34-41. https://doi.org/10.5688/aj690105

Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International education*, *5*(2), 131-154. https://doi.org/10.1177/1475240906065589

Annan, S. L., Tratnack, S., Rubenstein, C., Metzler-Sawin, E., & Hulton, L. (2013). An Integrative Review of Student Evaluations of Teaching: Implications for Evaluation of Nursing Faculty. *Journal of Professional Nursing, 29*(5), 10-24. doi:10.1016/j.profnurs.2013.06.004

Appling, S. E., Naumann, P. L., & Berk, R. A. (2001). Using a faculty evaluation triad to achieve evidence-based teaching. *Nursing Education Perspectives*, *22*(5), 247-251.

Armstrong, M. J. (2013). A Preliminary Study of Grade Forecasting by Students. *Decision Sciences Journal Of Innovative Education, 11*(2), 193-210. doi:10.1111/dsji.12003.

Arnold, I. J. M. (2009). Do examinations influence student evaluations? International Journal of Educational Research, *48*(4), 215-224. http://dx.doi.org/10.1016/j.ijer.2009.10.001

Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems*. Bolton, MA: Anker Publishing Company.

Aubrecht, J. D. (1984). Better faculty evaluation systems. In P. Seldin (Ed.), *Changing practices in faculty evaluation: A critical assessment and recommendations for improvement* (pp. 85-91). San Francisco, CA: Jossey-Bass.

Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *The Journal of Economic Education*, *37*(1), 21-37. https://doi.org/10.3200/JECE.37.1.21-37

Bain, K. (2011). *What the best college teachers do*. Cambridge MA: Harvard University Press.

Balam, E. M., & Shannon, D. M. (2010). Student ratings of college teaching: A comparison of faculty and their students. Assessment & Evaluation in Higher Education, *35*(2), 209–221. https://doi.org/10.1080/02602930902795901

Ballantyne, C. (2005, November). *Moving student evaluation of teaching online: reporting pilot outcomes and issues with a focus on how to increase student response rate*. Paper presented at the annual meeting the University of Sydney Learning and Teaching: Evaluating and Enhancing the Experience. Sydney, Australia. Retrieved from https://www.researchgate.net/publication/272831950_

Barth, M. (2008). Deciphering student evaluations of teaching: A factor analysis approach. *The Journal of Education for Business*, *84*(1), 40–46. https://doi.org/10.3200/JOEB.84.1.40-46

Basow, S. A. (1994). Student ratings of professors are not gender blind. *AWM Newsletter*, *24*(5), 1-14.

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently?. *Journal of educational psychology*, *79*(3), 308-314. https://doi.org/10.1037/0022-0663.79.3.308

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles?. *Psychological science in the public interest*, *4*(1), 1-44. https://doi.org/10.1111/1529-1006.01431

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review*, *31*(5), 709-719. https://doi.org/10.1016/j.econedurev.2012.05.001

Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, *74*(2), 170-179. https://doi.org/10.1037/0022-0663.74.2.170

Beran, T., & Violato, C. (2006). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education*, *30*(6), 593–601. https://doi.org/10.1080/02602930500260688

Beran, T., & Violato, C. (2009). Student ratings of teaching effectiveness: Student engagement and course characteristics. *The Canadian Journal of Higher Education*, *39*(1), 1–13. Retrieved from Eric database. (EJ849729)

Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A "Consequential validity" study. *The Canadian Journal of Higher Education*, *35*(2), 49–70. Retrieved from

http://www.internationalgme.org/Resources/Pubs/Beran%20et%20al%20-%20Consequential%20validity.pdf

Beran, T., Violato, C., Kline, D., & Frideres, J. (2009). What do students consider useful about student ratings? *Assessment & Evaluation in Higher Education*, *34*(5), 519–527. https://doi.org/10.1080/02602930802082228

Beran, T. N., & Rokosh, J. L. (2009a). The consequential validity of student ratings: What do instructors really think? *The Alberta Journal of Educational Research*, *55*(4), 497–511. Retrieved from Eric database. (EJ720732)

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

Beran, T. N., Violato, C., & Kline, D. (2007). What's the "use" of student ratings of instruction for administrators? One university's experience. *The Canadian Journal of Higher Education*, *37*(1), 27–43.

Berk, R. A. (1979). The construction of rating instruments for faculty evaluation: a review of methodological issues. *The Journal of Higher Education*, *50*(5), 650-669. https://doi.org/10.1080/00221546.1979.11779999

Betoret, F. D. (2007). The influence of students' and teachers' thinking styles on student course satisfaction and on their learning process. *Educational Psychology*, *27*(2), 219-234. https://doi.org/10.1080/01443410601066701

Blackhart, G. C., Peruche, B. M., DeWall, C. N., Joiner, Jr, T. E., Dickson, K. L., Miller, M. D., ... & Walter, M. I. (2006). Faculty forum. *Teaching of Psychology*, *33*(1), 37-63. https://doi.org/10.1207/s15328023top3301_9

Blunt, A. (1991). The effects of anonymity and manipulated grades on student ratings of instructors. *Community College Review*, *18*(4), 48-54. https://doi.org/10.1177/009155219101800409

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research, 1*(7), 1-11. Retrieved from https://www.scienceopen.com/document?vid=818d8ec0-5908-47d8-86b4-5dc38f04b23e

Bound, J., Braga, B., Khanna, G., & Turner, S. (2016, December). *A passage to America: University funding and international students*. Paper presented at the annual meeting of the National Bureau of Economic Research, Cambridge, MA. Retrieved from http://www.nber.org/papers/w22981.pdf

Brandenburg, D. C. (1979). Consideration for an Evaluation Program of Instructional Quality. *CEDR Quarterly*, *12*(4), 8-12.

Brandenburg, D. C., Slinde, J. A., & Batista, E. E. (1977). Student ratings of instruction: Validity and normative interpretations. *Research in Higher Education*, *7*(1), 67-78. https://doi.org/10.1007/BF00991945

Braskamp, L. A., & Ory, J. C. (1994). *Assessing Faculty Work: Enhancing Individual and Institutional Performance. Jossey-Bass Higher and Adult Education Series*. San Francisco, CA: Jossey-Bass Inc.

Brown, M. J., Baillie, M., & Fraser, S. (2009). Rating RateMyProfessors.com: A comparison of online and official student evaluations of teaching. *College Teaching*, *57*(2), 89-92. https://doi.org/10.3200/CTCH.57.2.89-92

Bruner, J. S. (2009). *The Process of Education*. Cambridge, MA: Harvard University Press.

Buck, S., & Tiene, D. (1989). The impact of physical attractiveness, gender, and teaching philosophy on teacher evaluations. *The Journal of Educational Research*, *82*(3), 172-177. https://doi.org/10.1080/00220671.1989.10885887

Campion, W. J., Mason, D. V., & Erdman, H. (2000). How faculty evaluations are used in Texas community colleges. *Community College Journal of Research & Practice*, *24*(3), 169-179. https://doi.org/10.1080/106689200264132

Cashin, W. E. (1988). *Student ratings of teaching: A summary of the research. IDEA Paper No. 20.* Retrieved from Eric database. (ED302567)

Cashin, W. E. (1990). Students do rate different academic fields differently. *New directions for teaching and learning*, *1990*(43), 113-121. https://doi.org/10.1002/tl.37219904310

Cashin, W. E. (1999). Student ratings of teaching: Uses and misuses. In P. Selden and Associates (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 25-44). Bolton, MA: Anker.

Cassidy, J. (1990). Effect of intensity training on preservice teachers' instruction accuracy and Delivery Effectiveness. *Journal of Research in Music Education*, *38*(3), 64-74. https://doi.org/10.2307/3345180

Catano, V. M., & Harvey, S. (2011). Student perception of teaching effectiveness: development and validation of the Evaluation of Teaching Competencies Scale (ETCS). *Assessment &*

Evaluation in Higher Education, *36*(6), 701-717. https://doi.org/10.1080/02602938.2010.484879

Centra, J. A. (1973). Self-ratings of college teachers: A comparison with student ratings. *Journal of Educational Measurement*, *10*(4), 287-295. https://doi.org/10.1111/j.1745-3984.1973.tb00806.x

Centra, J. A. (1974). The relationship between student and alumni ratings of teachers. *Educational and Psychological Measurement*, *34*(2), 321-325. https://doi.org/10.1177/001316447403400212

Centra, J. A. (1976). The influence of different directions on student ratings of instruction. *Journal of Educational Measurement*, *13*(4), 277-282. https://doi.org/10.1177/001316447403400212

Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. *American educational research journal*, *14*(1), 17-24. https://doi.org/10.3102/00028312014001017

Centra, J. A. (1983). Research productivity and teaching effectiveness. *Research in Higher education*, *18*(4), 379-389. https://doi.org/10.1007/BF00974804

Centra, J. A. (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness. The Jossey-Bass Higher and Adult Education Series*. San Francisco, CA: Jossey-Bass Inc.

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

Centra, J. A., & Creech, F. R. (1976). The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness. *Project report*, *76*(1), 1-59. Retrieved from https://www.ets.org/research/policy_research_reports/publications/report/1976/hylb

Chandler, T. A. (1978). The questionable status of student evaluations of teaching. *Teaching of Psychology*, *5*(3), 150-152. https://doi.org/10.1207/s15328023top0503_12

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. https://doi.org/10.1080/02602930301683

Chiang, H. H. (2005). Students' perceptions of effective and ineffective clinical instructors. *Journal of Nursing Education*, *44*(4), 187-192. Retrieved from NCBI database. (15862053)

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. https://doi.org/10.1177/0273475308324086

Clayson, D. E., & Haley, D. (2011). Are students telling us the truth? A critical look at the student evaluation of teaching. *Marketing Education Reviews*, *21*(2), 101–112. https://doi.org/10.2753/MER1052-8008210201

Clayson, D. E., & Sheffet, M. J. (2006). Personality and the student evaluation of teaching. *Journal of marketing education*, *28*(2), 149-160. https://doi.org/10.1177/0273475306288402

Cohen, C. E. (1981). Person Categories and Social Perception: Testing Some Boundaries of the Processing Effects of Prior Knowledge. J*ournal of Personality and Social Psychology*, *40*(3), 441-452. https://doi.org/10.1037/0022-3514.40.3.441

Cohen, J. (1995). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cranton, P. A., & Smith, R. A. (1986). A new look at the effect of course characteristics on student ratings of instruction. *American Educational Research Journal*, *23*(1), 117-128. https://doi.org/10.3102/00028312023001117

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. https://doi.org/10.1108/EUM0000000006158

Dommeyer, C. J., Baum, P., Chapman, K. S., & Hanna, R. W. (2002). Attitudes of business faculty towards two methods of collecting teaching evaluations: Paper vs. online. *Assessment & Evaluation in Higher Education*, *27*(5), 455-462. https://doi.org/10.1080/0260293022000009320

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. http://dx.doi.org/10.1080/02602930410001689171

Donnon, T., Delver, H., & Beran, T. (2010). Student and teaching characteristics related to ratings of instruction in medical sciences graduate programs. *Medical Teacher*, *32*(4), 327–332. https://doi.org/10.3109/01421590903480097

Donovan, J., Mader, C. E., & Shinsky, J. (2006). Constructive Student Feedback: Online vs. Traditional Course Evaluations. *Journal of Interactive Online Learning*, *5*(3), 283-296. Retrieved from https://www.ncolr.org/jiol/issues/pdf/5.3.5.pdf

Dowell, D. A., & Neal, J. A. (1982). A selective review of the validity of student ratings of teaching. *The Journal of Higher Education*, *53*(1), 51-62. https://doi.org/10.2307/1981907

Dudley, S., & Shawver, D. L. (1991). The effect of homework on students' perceptions of teaching effectiveness. *Journal of Education for Business*, *67*(1), 21-25. https://doi.org/10.1080/08832323.1991.10117511

Eble, K. E. (1970, November). *The recognition and evaluation of teaching*. Paper presented at the annual meeting of the American Association of University Professors, Washington D.C. Retrieved from Eric database. (ED046350)

El Hassan, K. (2009). Investigating substantive and consequential validity of student ratings of instruction. *Higher Education Research and Development*, *28*(3), 319–333. https://doi.org/10.1080/07294360902839917

Emerson, R. J., & Records, K. (2007). Design and testing of classroom and clinical teaching evaluation tools for nursing education. *International journal of nursing education scholarship*, *4*(1), 1-16. https://doi.org/10.2202/1548-923X.1375

Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education*, *4*(1), 69-111. https://doi.org/10.1007/BF00991462

Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education*, *6*(3), 223-274. https://doi.org/10.1007/BF00991288

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education*, *9*(3), 199-242. https://doi.org/10.1007/BF00976997

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, *10*(2), 149-172. https://doi.org/10.1007/BF00976227

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, *18*(1), 3-124. https://doi.org/10.1007/BF00992080

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education*, *21*(1), 45-116. https://doi.org/10.1007/BF00975035

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in higher education*, *24*(2), 139-213. https://doi.org/10.1007/BF00991885

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II—Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, *34*(2), 151-211. https://doi.org/10.1007/BF00992161

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-143). New York, NY: Springer. https://doi.org/10.1007/1-4020-5742-3_5

Fiske Jr, H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education*, *25*(4), 256-263. https://doi.org/10.2307/3345266

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972-1977*. New York, NY: Pantheon Books.

Frankhouser, W. M. (1984). The effects of different oral directions as to disposition of results on student ratings of college instruction. *Research in Higher Education*, *20*(3), 367-374. https://doi.org/10.1007/BF00983509

Franklin, J., & Theall, M. (1989, March). *Who Reads Ratings: Knowledge, Attitude, and Practice of Users of Student Ratings of Instruction*. Paper presented at the annual meeting of the American Education Research Association. San Francisco, CA. Retrieved from Eric database. (ED306241)

Franklin, J. L., & Theall, M. (1991, April). *Grade inflation and student ratings: A closer look*. Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL. Retrieved from Eric database. (ED349318)

Fraser, B. J., & Treagust, D. F. (1986). Validity and use of an instrument for assessing classroom psychosocial environment in higher education. *Higher education*, *15*(1-2), 37-57. https://doi.org/10.1007/BF00138091

Frey, P. W. (1976). Validity of student instructional ratings: Does timing matter?. *The Journal of Higher Education*, *47*(3), 327-336. https://doi.org/10.1080/00221546.1976.11774047

Frick, T. W., Chadha, R., Watson, C., & Zlatkovska, E. (2010). Improving course evaluations to improve instruction and complex learning in higher education. *Educational Technology Research and Development*, *58*(2), 115-136. Retrieved from https://www.jstor.org/stable/40603152

Gall, G., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction*, *7*. Boston, MA: Pearson/Allyn & Bacon.

Gallagher, T. (2000). Embracing Student Evaluations of Teaching: A Case Study. *Teaching Sociology, 28*(2), 140-147. https://doi.org/10.2307/1319261

Goldman, L. (1993). On the erosion of education and the eroding foundations of teacher education (or why we should not take student evaluation of faculty seriously). *Teacher Education Quarterly*, *20*(2), 57-64.

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. https://doi.org/10.1080/02602938.2014.960364

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, *52*(11), 1182-1186. https://doi.org/10.1037/0003-066X.52.11.1182

Guder, F., & Malliaris, M. (2010). Online and paper course evaluations. *American Journal of Business Education*, *3*(2), 131–137. https://doi.org/10.19030/ajbe.v3i2.392

Gump, S. (2007). Student evaluations of teaching effectiveness and the leniency hypothesis: A literature review. *Educational Research Quarterly*, *30*(3), 55–68. Retrieved from Eric database. (EJ787711)

Guthrie, E. R. (1954). *The evaluation of teaching: A progress report*. Seattle, WA: University of Washington.

Hajdin, G., & Pažur, K. (2012). Differentiating between student evaluation of teacher and teaching effectiveness. *Journal of Information and Organizational Sciences*, *36*(2), 123-134. Retrieved from https://hrcak.srce.hr/93738

Haskell, R. E. (1997). Administrative use of student evaluation of faculty. *Education policy analysis archives*, *5*(21), 1-44. https://doi.org/10.14507/epaa.v5n21.1997

Hativa, N. (1995). The department-wide approach to improving faculty instruction in higher education: A qualitative evaluation. *Research in Higher Education*, *36*(4), 377-413. https://doi.org/10.1007/BF02207904

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), 1-9. Retrieved from Eric database. (EJ765357)

Heller, S. (1984). Math department balks at officials' effort to require forms for student evaluation. *Chronicle of Higher Education*, *28*(8), 24.

Hill, M. C., & Epps, K. K. (2010). The impact of physical classroom environment on student satisfaction and student evaluation of teaching in the university environment. *Academy of Educational Leadership Journal*, *14*, 65–79.

Hills, S. B., Naegle, N., & Bartkus, K. R. (2009). How important are items on a student evaluation? A study of item salience. *The Journal of Education for Business*, *84*(5), 297–303. https://doi.org/10.3200/JOEB.84.5.297-303

Hilt, D. (2001). What Students Can Teach Professors: Reading Between the Lines of Evaluations. *Chronicle of Higher Education*, *47*(27), 15-25.

Hmieleski, K. & Champagne, M. V. (2000). *Plugging in to course evaluation*. Retrieved from http://technologysource.org/article/plugging_in_to_course_evaluation

Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations. *College Teaching, 49*(1), 26-31. https://doi.org/10.1080/87567550109595842

Hofman, J. E., & Kremer, L. (1980). Attitudes toward higher education and course evaluation. *Journal of Educational Psychology*, *72*(5), 610-617. https://doi.org/10.1037/0022-0663.72.5.610

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. *Journal of Educational Psychology*, *77*(2), 187-196. https://doi.org/10.1037/0022-0663.77.2.187

Huff, D. (1954). *How to Lie with Statistics*. New York, NY: W.W. Norton & Company Inc.

Hussar, W. J. and Bailey, T. M. (2017). *Projections of Education Statistics to 2025*. Retrieved from https://nces.ed.gov/pubs2017/2017019.pdf

Johnson, D. I. (2009). Connected classroom climate: A validity study. *Communication Research Reports*, *26*(2), 146-157. https://doi.org/10.1080/08824090902861622

Jonassen, D., & Land, S. (2012). *Theoretical foundations of learning environments* (2nd ed.). New York, NY: Routledge

Kaiser, K. A. (1998). *The effect of differentiated high- versus low-intensity teaching on band musicians' evaluation of teaching effectiveness* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 9827669).

Kaschak, E. (1978). Sex bias in student evaluations of college professors. *Psychology of Women Quarterly*, *2*(3), 235-243. https://doi.org/10.1111/j.1471-6402.1978.tb00505.x

Kasiar, J. B., Schroeder, S. L., & Holstad, S. G. (2002). Comparison of traditional and web-based course evaluation processes in a required, team-taught pharmacotherapy course. *American Journal of Pharmaceutical Education*, *66*(3), 268-270.

Kember, D., Leung, D. Y., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching?. *Assessment & Evaluation in Higher Education*, *27*(5), 411-425. https://doi.org/10.1080/0260293022000009294

Kiefer, K. (1994). *Academic Evaluations*. Retrieved from https://wac.colostate.edu/resources/writing/guides/evaluations/

Koblitz, N. (1990). Are student ratings unfair to women?, *Newsletter of the Association for Women in Mathematics*, *20*(1), 17-19.

Koermer, C. D., & Petelle, J. L. (1991). Expectancy violation and student rating of instruction. *Communication Quarterly*, *39*(4), 341-350. https://doi.org/10.1080/01463379109369810

Kolitch, E., & Dean, A. V. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about ‘good’ teaching. *Studies in Higher Education, 24*(1), 27-42. https://doi.org/10.1080/03075079912331380128

Koon, J., & Murray, H. G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. *The Journal of Higher Education*, *66*(1), 61-81. https://doi.org/10.2307/2943951

Koushki, P. A., & Kuhn, H. A. J. (1982). How Reliable are Student Evaluations of Teachers. *Engineering Education*, *72*(5), 362-367.

Kress. G. (2000). Multimodality. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 182–202). London, England: Routlege

Kwan, K. P. (1999). How fair are student ratings in assessing the teaching performance of university teachers?. *Assessment & Evaluation in Higher Education*, *24*(2), 181-195. https://doi.org/10.1080/0260293990240207

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. https://doi.org/10.1016/j.econedurev.2006.12.003

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.

Layne, B. H., DeCristoforo, J. R., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, *40*(2), 221-232. https://doi.org/10.1023/A:1018738731032

Leventhal, L., Abrami, P. C., & Perry, R. P. (1976). Do teacher rating forms reveal as much about students as about teachers?. *Journal of Educational Psychology*, *68*(4), 441-445. https://doi.org/10.1037/0022-0663.68.4.441

L'hommedieu, R., Menges, R. J., & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology*, *82*(2), 232-241. https://doi.org/10.1037/0022-0663.82.2.232

Macfadyen, L. P., Dawson, S., Prest, S., & Gašević, D. (2016). Whose feedback? A multilevel analysis of student completion of end-of-term teaching evaluations. *Assessment & Evaluation in Higher Education*, *41*(6), 821-839. https://doi.org/10.1080/02602938.2015.1044421

Mansson, D. H. (2014). Students' Expressed Academic Concern, Learning Outcomes, and Communication Motives. *Western Journal of Communication, 78*(3), 274-286. http://dx.doi.org/10.1080/10570314.2014.904521

March, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. *Higher education: Handbook of theory and research*, *8*(1), 143-233.

Marks, R. B. (2000). Determinants of student evaluations of global measures of instructor and course value. *Journal of Marketing Education*, *22*(2), 108-119. https://doi.org/10.1080/10570314.2014.904521

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, *76*(5), 707-754. https://doi.org/10.1037/0022-0663.76.5.707

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International journal of educational research*, *11*(3), 253-388. https://doi.org/10.1016/0883-0355(87)90001-2

Marsh. H.W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319–383). Dordrecht, The Netherlands: Springer.

Marsh, H. W., & Cooper, T. L. (1981). Prior subject interest, students' evaluations, and instructional effectiveness. *Multivariate Behavioral Research*, *16*(1), 83-104. https://doi.org/10.1207/s15327906mbr1601_5

Marsh, H. W., & Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology*, *72*(4), 468-475. https://doi.org/10.1037/0022-0663.72.4.468

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American educational research journal*, *30*(1), 217-251. https://doi.org/10.3102/00028312030001217

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders?. *Journal of Educational Psychology*, *92*(1), 202-228. https://doi.org/10.1037/0022-0663.92.1.202

Marsh, H. W., & Ware, J. E. (1982). Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr. Fox effect. *Journal of educational Psychology*, *74*(1), 126-134. https://doi.org/10.1037/0022-0663.74.1.126

Matos-Díaz, H., & Ragan Jr, J. F. (2010). Do student evaluations of teaching depend on the distribution of expected grade?. *Education Economics*, *18*(3), 317-330. https://doi.org/10.1080/09645290903109444

McCallum, L. W. (1984). A meta-analysis of course evaluation data and its use in the tenure decision. *Research in Higher Education*, *21*(2), 150-158. https://doi.org/10.1007/BF00975102

McCroskey, J. C. (1994, April). *Assessment of affect toward communication and affect toward instruction in communication*. Paper presented at the annual meeting of the Speech Communication Association, Annandale, VA.

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, *65*(6), 384-397. https://doi.org/10.2307/40248725

McKeachie, W. J. (1990). Research on college teaching: The historical background. *Journal of educational psychology*, *82*(2), 189-200. https://doi.org/10.1037/0022-0663.82.2.189

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*(11), 1218-1225. http://dx.doi.org/10.1037/0003-066X.52.11.1218

McKeachie, W. J. (2002). *Teaching tips: Strategies, research, and theory for college and university professors.* Orlando, FL: Houghton-Mifflin Publishing.

McNulty, J. A., Gruener, G., Chandrasekhar, A., Espiritu, B., Hoyt, A., & Ensminger, D. (2010). Are online student evaluations of faculty influenced by the timing of evaluations?. *Advances in physiology education*, *34*(4), 213-216. https://doi.org/10.1152/advan.00079.2010

Menefee, R. (1983). The evaluation of science teaching. *Journal of College Science Teaching*, *28*(1), 138.

Miller, A. H. (1988). Student assessment of teaching in higher education. *Higher Education*, *17*(1), 3-15. https://doi.org/10.1007/BF00130896

Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of educational psychology*, *82*(2), 250-261. https://doi.org/10.1037/0022-0663.82.2.250

Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of medical education*, *48*(7), 630-635. Retrieved from http://adrianmarriott.net/logosroot/papers/DrFoxSpoof.pdf

National Association of Schools of Music. (2016). *Purposes*. Retrieved from https://nasm.arts-accredit.org/about/purposes/

Nonaka, I., & Takeuchi, H. (1996). The knowledge-creating company: How Japanese companies create the dynamics of innovation. *Long range planning*, *4*(29), 592-611. https://doi.org/10.1016/0024-6301(96)81509-3

Nowell, C., & Alston, R. M. (2007). I thought I got an A! Overconfidence across the economics curriculum. *The Journal of Economic Education*, *38*(2), 131-142. https://doi.org/10.3200/JECE.38.2.131-142

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. https://doi.org/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. https://doi.org/10.1080/02602930701293231

Olshavsky, R., & Spreng, R. A. (1995). Consumer satisfaction and students: Some pitfalls of being customer driven. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *8*(2), 69-77.

Oppenheim, A.(1992). Questionnaire Design, Interviewing and Attitude Measurement. *Journal of Community & Applied Social Psychology*, *4*(5), 371-372. https://doi.org/10.1002/casp.2450040506

Ory, J. C. (2000). Teaching evaluation: Past, present, and future. *New directions for teaching and learning*, *2000*(83), 13-18. https://doi.org/10.1002/tl.8302

Ory, J. C., & Braskamp, L. A. (1981). Faculty perceptions of the quality and usefulness of three types of evaluative information. *Research in Higher Education*, *15*(3), 271-282. https://doi.org/10.1007/BF00976421

Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework?. *New directions for institutional research*, *2001*(109), 27-44. https://doi.org/10.1002/ir.2

Overall, J. U., & Marsh, H. W. (1979). Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. *Journal of Educational Psychology*, *71*(6), 856-865. https://doi.org/10.1037/0022-0663.71.6.856

Page, C. F. (1974). *Student evaluation of teaching: the American experience*, *22*. London, England: Society for Research into Higher Education.

Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in higher education*, *8*(3), 399-411. https://doi.org/10.1080/13562510309396

Perry, R. P. (1990). Introduction to the special section. *Journal of Educational Psychology*, *82*(2), 183-188. https://doi.org/10.1037/0022-0663.82.2.183

Perry, R. P., Abrami, P. C., Leventhal, L., & Check, J. (1979). Instructor reputation: An expectancy relationship involving student ratings and achievement. *Journal of Educational Psychology*, *71*(6), 776-787. https://doi.org/10.1037/0022-0663.71.6.776

Perry, R. P., Niemi, R. R., & Jones, K. (1974). Effect of prior teaching evaluations and lecture presentation on ratings of teaching performance. *Journal of Educational Psychology*, *66*(6), 851-856. https://doi.org/10.1037/h0021527

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education*, *15*(2), 178-191. https://doi.org/10.1108/09684880710748938

Powell, R. W. (1977). Grades, learning, and student evaluation of instruction. *Research in Higher Education*, *7*(3), 193-205. https://doi.org/10.1007/BF00991986

Prave, R. S., & Baril, G. L. (1993). Instructor ratings: Controlling for bias from initial student interest. *Journal of Education for Business*, *68*(6), 362-366. https://doi.org/10.1080/08832323.1993.10117644

United States Department of Education. (2014). *Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices*. Retrieved from

https://studentprivacy.ed.gov/resources/protecting-student-privacy-while-using-online-educational-services-requirements-and-best

Pulich, M. A. (1984). Ratings: Better Use of Student Evaluations for Teaching Effectiveness. *Improving College and University Teaching*, *32*(2), 91-94. https://doi.org/10.1080/00193089.1984.10533850

Quinn, D. (2002). *Improving online response rates*. Retrieved from http://www.unisanet.unisa.edu.au/sei/website/Online-respnrates.asp

Ragan Jr, J. F., & Walia, B. (2010). Differences in student evaluations of principles and other economics courses and the allocation of faculty across courses. *The Journal of Economic Education*, *41*(4), 335-352. https://doi.org/10.1080/00220485.2010.510389

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in higher education*, *16*(2), 129-150. https://doi.org/10.1080/03075079112331382944

Remmers, H. H. (1928). The relationship between students' marks and student attitude toward instructors. *School & Society, 28,* 759-760.

Remmers, H. H., & Brandenburg, G. C. (1927). Experimental data on the Purdue rating scale for instructors. *Educational Administration and Supervision*, *13*(6), 399-406.

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

Rubin, H. J., & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage Publications.

Rustagi, N. K. (1997). A study of the retention of basic quantitative skills. *Journal of Education for Business*, *73*(2), 72-76. https://doi.org/10.1080/08832329709601619

Ryan, J. J., Anderson, J. A., & Birchler, A. B. (1980). Student evaluation: The faculty responds. *Research in Higher Education*, *12*(4), 317-333. https://doi.org/10.1007/BF00976185

Scherr, F. C., & Scherr, S. S. (1990). Bias in Student Evaluations of Teacher Effectiveness. *Journal of Education for Business*, *65*(8), 356-58.

Schoofs, N. C. (1997). *A conceptual analysis of the views of student evaluations of faculty at three liberal arts colleges* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 9731868)

Schulz, F. (2013). *Distance Music. From early telephony to tomorrow's Internet* (Unpublished doctoral dissertation). University of Auckland, New Zealand.

Scott, C. S. (1977). Student ratings and instructor-defined extenuating circumstances. *Journal of Educational Psychology*, *69*(6), 744-747. https://doi.org/10.1037/0022-0663.69.6.744

Scriven, M. (1981). Summative Teacher Evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation: Assessing elementary and secondary school teachers* (pp.244-271). Newbury Park, CA: Sage Publications.

Seldin, P. (1984). *Changing practices in faculty evaluation*. San Francisco, CA: Jossey-Bass, Inc., Publications.

Seldin, P. (1993). The use and abuse of student ratings of professors. *The chronicle of higher Education*, *39*(46), 40.

Seldin, P. (1999). Building successful teaching evaluation programs. In P. Seldin (Ed.), *Current practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 240-242). Bolton, MA: Anker.

Serdyukova, N., Tatum, B. C., & Serdyukov, P. (2010). Student evaluations of courses and teachers. *Journal of Research in Innovative Teaching*, *3*(1), 180–189. Retrieved from https://www.nu.edu/assets/resources/pageResources/journal-of-research-in-innovative-teaching-volume-3.pdf

Sheehan, D. S. (1975). On the invalidity of student ratings for administrative personnel decisions. *The Journal of Higher Education*, *46*(6), 687-700. https://doi.org/10.2307/1979062

Sheets, D. F., Topping, E. E., & Hoftyzer, J. (1995). The relationship of student evaluations of faculty to student performance on a common final examination in the principles of economics course. *The Journal of Economics*, *21*(2), 55-64.

Shih, C. L., & Chuang, H. H. (2013). The development and validation of an instrument for assessing college students' perceptions of faculty knowledge in technology-supported class environments. *Computers & Education*, *63*, 109-118. https://doi.org/10.1016/j.compedu.2012.11.021

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. https://doi.org/10.1177/0273475300223004

Small, A. C., Hollenbeck, A. R., & Haley, R. L. (1982). The effect of emotional state on student ratings of instructors. *Teaching of Psychology*, *9*(4), 205-211. https://doi.org/10.1207/s15328023top0904_3

Sojka, J., Gupta, A. K., & Deeter-Schmelz, D. R. (2002). Student and faculty perceptions of student evaluations of teaching: A study of similarities and differences. *College Teaching*, *50*(2), 44-49. https://doi.org/10.1080/87567550209595873

Sonntag, M. E., Bassett, J. F., & Snyder, T. (2009). An empirical test of the validity of student evaluations of teaching made on RateMyProfessors.com. *Assessment & Evaluation in Higher Education*, *34*(5), 499-504. https://doi.org/10.1080/02602930802079463

Spencer, P.A., and Flyr, M.L. (1992). *The formal evaluation as an impetus to classroom change: Myth or reality?* Riverside, CA: University of California Press.

Sprinkle, J. E. (2008). Student perceptions of effectiveness: An examination of the influence of student biases. *College Student Journal*, *42*(2), 276-293.

Stark, P., & Freishtat, R. (2014). An Evaluation of Course Evaluations. *ScienceOpen Research*, *1*(2), 1-7. https://doi.org/10.14293/S2199-1006.1.sor-edu.aofrqa.v1

Tagomori, H. T., & Bishop, L. A. (1995). Student Evaluation of Teaching: Flaws in the Instruments. *Thought & Action*, *11*(1), 63-78.

Tatro, C.N. (1995). Gender Effects of Student Evaluations of Faculty. *Journal of Research and Development in Education, 28*(1), 169–173.

Theall, M., & Franklin, J. (1991). Using student ratings for teaching improvement. *New directions for teaching and learning*, *1991*(48), 83-96. https://doi.org/10.1002/tl.37219914808

Thompson Jr, R. J., & Serra, M. (2005). Use of Course Evaluations to Assess the Contributions of Curricular and Pedagogical Initiatives to Undergraduate General Education Learning Objectives. *Education*, *125*(4), 693-701.

Thorpe, S. W. (2002, June). *Online student evaluation of instruction: An investigation of non-response bias*. Paper presented at the Annual Forum of the Association for Institutional Research. Toronto, Ontario, Canada. Retrieved from http://www.memphis.edu/sete/pdfs/online-eval.pdf

Tollefson, N., Chen, J. S., & Kleinsasser, A. (1989). The relationship of students' attitudes about effective teaching to students' ratings of effective teaching. *Educational and Psychological Measurement*, *49*(3), 529-536. https://doi.org/10.1177/001316448904900303

Vaughan, C. J. (2011). *An investigation of band directors' perceptions of importance of the most common student teaching learning outcomes* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 3586343).

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

Wallace, J. J., & Wallace, W. A. (1998). Why the costs of student evaluations have long since exceeded their value. *Issues in Accounting Education*, *13*(2), 443-448.

Watt, S., Simpson, C., McKillop, C., & Nunn, V. (2002). Electronic course surveys: does automating feedback and reporting give better results?. *Assessment & Evaluation in Higher Education*, *27*(4), 325-337. https://doi.org/10.1080/0260293022000001346

Waxman, H. C., & Huang, S. Y. L. (1997). Classroom instruction and learning environment differences between effective and ineffective urban elementary schools for African American students. *Urban Education*, *32*(1), 7-44. https://doi.org/10.1177/0042085997032001002

Whitworth, J. E., Price, B. A., & Randall, C. H. (2002). Factors that affect college of business student opinion of teaching and learning. *Journal of Education for Business*, *77*(5), 282-289. https://doi.org/10.1080/08832320209599677

Wicks, D. (2004). The institution of tenure: freedom or discipline?. *Management Decision*, *42*(5), 619-627. https://doi.org/10.1108/00251740410538479

Wilhelm, W. B. (2004). The relative influence of published teaching evaluations and other instructor attributes on course choice. *Journal of Marketing Education*, *26*(1), 17-30. https://doi.org/10.1177/0273475303258276

Wilhelm, W. B., & Comegys, C. (2004). Course selection decisions by students on campuses with and without published teaching evaluations. *Practical assessment, research & evaluation*, *9*(16), 1-20.

Wilson, R. (1998). *New Research Casts Doubt on Value of Student Evaluations of Professors*. Retrieved from https://www.chronicle.com/article/New-Research-Casts-Doubt-on/101118

Witt, J., & Burdalski, M. (2003, September). *Regarding the Timing of Student Course/Instructor Evaluations*. Paper presented at the annual meeting of the American Business Educators Conference, San Francisco, CA. Retrieved from https://bit.ly/2Ps7WrN

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, *17*(2), 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Woodward, D. K. (1998). Comparison of course evaluations by traditional and computerized online methods. *American Journal of Pharmaceutical Education*, *62*(1), 90.

Wright, R. E. (2006). Student evaluations of faculty: Concerns raised in the literature, and possible solutions. *College Student Journal*, *40*(2), 417-422.

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

Zhou, Y. (2018). *Foreign students continue to turn away from US universities*. Retrieved from https://qz.com/1267351/f-1-and-m-1-visa-data-show-international-students-are-turning-away-from-us-universities/

Zoller, U. (1992). Faculty Teaching Performance Evaluation in Higher Science Education: Issues and Implications (A "Cross-Cultural" Case Study). *Science Education*, *76*(6), 673-684. https://doi.org/10.1002/sce.3730760607

Zúñiga, R. E. (2004). Increasing response rates for online surveys—a report from the Flashlight Program's BeTA Project. *News for the flashlight program*, *4*(3) 14-18.

APPENDIX A

IRB APPROVAL LETTER

PEPPERDINE

Pepperdine University
24255 Pacific Coast Highway
Malibu, CA 90263
TEL: 310-506-4000

NOTICE OF APPROVAL FOR HUMAN RESEARCH

Date: March 18, 2018

Protocol Investigator Name: Donald Gruendler

Protocol #: 17-11-662

Project Title: STUDENT EVALUATIONS OF TEACHING: PERCEPTIONS OF FACULTY KNOWLEDGE AND THEIR RELATION TO LEARNING

School: Graduate School of Education and Psychology

Dear Donald Gruendler:

Thank you for submitting your application for exempt review to Pepperdine University's Institutional Review Board (IRB). We appreciate the work you have done on your proposal. The IRB has reviewed your submitted IRB application and all ancillary materials. Upon review, the IRB has determined that the above entitled project meets the requirements for exemption under the federal regulations 45 CFR 46.101 that govern the protections of human subjects.

Your research must be conducted according to the proposal that was submitted to the IRB. If changes to the approved protocol occur, a revised protocol must be reviewed and approved by the IRB before implementation. For any proposed changes in your research protocol, please submit an amendment to the IRB. Since your study falls under exemption, there is no requirement for continuing IRB review of your project. Please be aware that changes to your protocol may prevent the research from qualifying for exemption from 45 CFR 46.101 and require submission of a new IRB application or other materials to the IRB.

A goal of the IRB is to prevent negative occurrences during any research study. However, despite the best intent, unforeseen circumstances or events may arise during the research. If an unexpected situation or adverse event happens during your investigation, please notify the IRB as soon as possible. We will ask for a complete written explanation of the event and your written response. Other actions also may be required depending on the nature of the event. Details regarding the timeframe in which adverse events must be reported to the IRB and documenting the adverse event can be found in the *Pepperdine University Protection of Human Participants in Research: Policies and Procedures Manual* at community.pepperdine.edu/irb.

Please refer to the protocol number denoted above in all communication or correspondence related to your application and this approval. Should you have additional questions or require clarification of the contents of this letter, please contact the IRB Office. On behalf of the IRB, I wish you success in this scholarly pursuit.

Sincerely,

Judy Ho, Ph.D., IRB Chair

cc: Dr. Lee Kats, Vice Provost for Research and Strategic Initiatives

Mr. Brett Leach, Regulatory Affairs Specialist