The Effect of Student Evaluations on Faculty Performance

A Doctoral Thesis by

Kenneth M. Hartford

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Education

In Higher Education Administration

Northeastern University

February 2017
# TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................... 3

LIST OF TABLES:

Table 1: Pseudonyms for Study Participants ......................................................................................... 42
Table 2: Superordinate Themes and Subthemes ...................................................................................... 52
Figure 1: Consequences of Student Evaluations ................................................................................. 101

ACKNOWLEDGEMENTS .................................................................................................................. 4

1.  INTRODUCTION .......................................................................................................................... 5
2.  REVIEW OF LITERATURE .............................................................................................................. 17
3.  RESEARCH DESIGN ....................................................................................................................... 38
4.  RESEARCH FINDINGS ................................................................................................................... 51
5.  DISCUSSION AND IMPLICATIONS FOR PRACTICE .................................................................. 86

REFERENCES ..................................................................................................................................... 109

SCHEDULES:

A.  INVITATION TO PARTICIPATE ....................................................................................................... 126
B.  INFORMED CONSENT FORM ........................................................................................................ 127
C.  LETTER OF APPROVAL FROM INSTITUTIONAL REVIEW BOARD .............................................. 128
D.  INTERVIEW PROTOCOL FORM .................................................................................................... 129
E.  INTERVIEW # 1 FORM ................................................................................................................... 130
F.  INTERVIEW # 2 FORM ................................................................................................................... 132
ABSTRACT

THE EFFECT OF STUDENT EVALUATIONS ON FACULTY PERFORMANCE

By Ken Hartford

The purpose of this study was to understand the behavior of professors as a consequence of receiving student evaluations. To address this purpose, a single research question directed this study: In what ways and to what extent might faculty compromise good pedagogical practices for positive student evaluations of teaching?

A qualitative research design was used to understand the phenomenon of student evaluations. The study was a descriptive phenomenology because it describes an experience instead of interpreting it.

Data was collected through semi-structured, in-depth interviews, document and literature review, and analysis of field notes and journals. Participants for this study were 10 full-time professors from a college in Michigan. The researcher was the primary instrument in data collection and analysis. Data analysis went on concurrently with data collection. This gave the researcher an opportunity to test emerging themes and categories against subsequent data. Internal validity and external validity was provided. Reliability was also addressed. All data was kept confidential and stored in a secure location.

Key Words: evaluations, qualitative, phenomenology, analysis, validity
ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and guidance of a number of exceptional people.

I extend a heartfelt thank you to my academic advisor Dr. Kimberly Nolan. Her timely assistance and guidance was always appreciated during a sometimes long and arduous process. She was dedicated to my success and I am really grateful for that.

I want to thank Dr. Joseph McNabb for his input and direction. He is a true teacher and has a great passion for education. I am privileged to have crossed paths with him.

I want to express my gratitude to Dr. Tom Williams for his service on my dissertation committee. He is an experienced and enthusiastic educator and I appreciate his contribution and assistance.

I am fortunate to have a supportive family. My father Murray and sisters Diane and Vicky always had encouraging words during the thesis process. Son Kevin, the real writer in the family, always had something positive to contribute.

My biggest cheerleader was my wife Elizabeth. Her constant optimism and positive approach to life truly amazes me. She accompanied me for every step of this journey. I thank her from the bottom of my heart.
Chapter 1: Introduction to the Study

The purpose of this study was to determine the effect of student evaluations on faculty classroom pedagogy. The investigation determined if faculty compromise good teaching pedagogy in order to receive favorable ratings from students. The information from this study informs students, faculty, and administrators about student evaluations of faculty and their effect on teaching.

Context and Background

Most colleges and universities evaluate their faculty to determine if their work is satisfactory. Chairs, deans, and provosts need valid information to determine if professors are meeting performance standards as educators (Fink, 2008). Their classroom management, their teaching methods, and a number of important criteria should be evaluated to determine if they are performing satisfactorily and to ascertain if the students are being well served. Student evaluations are the primary tool for measuring the teaching effectiveness of professors in most colleges and universities (Huemer, 1998). Clayson and Haley (2011) found that almost one hundred percent of business schools in the United States utilize student ratings of faculty performance. Faculty can take the formative results and use them to improve their teaching. Administrative staff can take the summative results and make some important human resource decisions.

Formal faculty evaluations have been around for a long time. The earliest documented use of student evaluations was at the University of Washington in the 1920s (Murray, 2005). Student evaluations began in earnest in the late 1960s and early 1970s (Murray, 2005).

The initial aim of faculty evaluations in the 1960s was twofold: provide feedback to professors to help them improve their teaching and to assess the quality of teaching in
universities and colleges (Berk, 2005). Evaluations were conducted because students wanted a say in their instruction. They wanted to assess the effectiveness of their professors’ teaching methods. Administrators wanted to demonstrate that they were proactive about teacher quality. The same holds true today.

Some researchers have concluded that student evaluations do not truly measure teaching performance. Some suggest that students are biased and are not trained to rate teaching skills (winship, 2011). winship has also posited that evaluations merely measure student satisfaction and little else. Langbein (2008) posits that students aren’t proficient in evaluating faculty performance and the results are problematic.

Other researchers have concluded that student evaluations contribute to grade inflation (Germain & Scandura, 2005). Arum and roksa (2010) contend that very little undergraduate learning takes place on today’s college campuses and student evaluations are a primary cause. They suggest that part of the reason is that student evaluations are the principal determinant of faculty performance and faculty have learned how to manipulate the ratings in their favor by reducing course rigor and demanding little from students.

An important consideration regarding faculty evaluations is whether or not student evaluation forms are reliable and valid. Reliability refers to congruity across raters, across time and across items (Ali & Sell, 1998). Reliability can refer to test-retest consistency where responses are constant across time. It can also apply to inter-rater reliability where many students give similar ratings to a professor. Reliability can also pertain to internal consistency where an instructor gets similar marks in different categories. Reliability refers to how consistent an instrument measures a set of constructs at different times and in different settings.
Concerns about reliability were highest when rating forms were constructed in-house and not prepared professionally or subject to psychometric testing (Arreola, 1995). Aleamoni (1987) and Arreola (1995) conducted literature reviews about student rating forms and concluded that tested and well-designed forms were usually considered reliable and valid.

Validity in professor evaluations is the degree to which a rating form measures what it was designed to measure. The existing evidence supports the validity of faculty evaluations when they are compared with other assessment methods like trained observation, peer appraisals, and measures of student learning (Arreola, 1995). Aleamoni and Hexner (1980) reviewed 14 studies in which student evaluations were compared to other methods of assessing faculty teaching and moderate to high correlations were found. The strong relationship between student evaluations and other kinds of faculty evaluation were supported by similar studies analyzed by Murray (1984) and McKeachie (1979). They concluded that student evaluations of faculty performance corresponded to the more neutral and expert assessment of colleagues, peers, and trained observers (Murray, 1984; McKeachie, 1979).

Greenwald (1997) reported that from 1971 to 1995 a majority of expert opinions confirmed the validity of student evaluations. Odden, Borman, and Fermanich (2004) found that research results suggest that educators have learned how to design and operate performance-based teacher assessments that are reliable and valid to use for consequential decisions. There is more to consider besides the reliability and validity of faculty evaluations.

The problem of practice is the purported connection between student evaluations of teaching and faculty pedagogical practices. Faculty may compromise their sound teaching methods to get better student evaluations. This is a major dilemma for academia. Therefore, this
study examined how participants discern these assessments and whether their pedagogy changes in order to gain a favorable student rating.

**Rationale and Significance**

The rationale for this study was to develop a further understanding of the phenomena of student evaluations and how they may affect faculty pedagogy. Research has suggested that faculty have reduced the rigor of courses and made tests easier in order to get positive evaluation results (Rice, 1988). A positive student evaluation may lead to faculty rewards.

The connection between student evaluations of faculty and purported grade inflation is significant. Well-educated citizens are required for today’s global economy. Inflated grades mean that students graduate without the requisite skills required by employers. The value of education will subsequently diminish and the importance of a degree will decline. Education is the foundation for a nation’s economic growth and prosperity. Anything that affects the education dynamic must be examined and related problems must be solved.

Student evaluations are often the only measure of faculty performance and administrators rely on this information regarding important human resource decisions. Subsequent personnel actions may include offering a professor tenure, awarding raises or promotions, and professional development opportunities. Decisions about programs and budget funding may also be contingent upon the results of student evaluations.

Researchers have suggested that there are problems with every facet of student evaluations. Huemer (1998) posited that students are not trained to evaluate good teaching. He also reported that students are biased when they rate their professors. He suggested that student ratings are influenced by superficial factors like gender, age, attractiveness, and lecture style.
Huemer suggests that faculty should not be judged based on student happiness. The student-as-customer metaphor exists at many colleges and universities. Schools have to shift their focus from delivering student happiness to producing quality education for their students. Some like Fink (2008) suggest that professors should be evaluated by their peers or they could self-report on a number of criteria like course design, professor-student interaction, quality of student learning, and ways one tried to improve their teaching over time.

Another practical goal is to re-examine the meaning of effective teaching. Joyce and Weil (1996) concluded that there is no single correct way to teach. This finding leads one to believe that good teaching has many facets and consequently a new, broader definition of good instruction should be developed. Researchers like Fink (2008), Huemer (1998), and Winship (2011) argue that students should not be the sole determinates of effective teaching.

A final practical goal is the examination of grade inflation and determining if student evaluations are one of the main reasons for grade inflation. Some studies suggest that the current process of students evaluating faculty contributes to grade inflation (Germain & Scandura, 2005). Fink (2008) articulated in a paper that one way to eliminate grade inflation is to discontinue student evaluations. He suggests that peer evaluation and self-evaluation are superior to the current method. Fink (2008) suggests that administrators should not transfer the responsibility for faculty assessment to students.

An intellectual goal of this study is understanding the process of student evaluations of faculty. Student evaluations must be valid. One should examine every aspect of the process to see if there are problems with the design of the ratings instrument, the timing of evaluations, the evaluation questions, and the context. It is important to determine the validity of the process
because if the design and process are satisfactory then one must conclude that any problems
created by evaluations must have another source.

This study is significant because education is the foundation for society. Faculty that
compromise their pedagogy diminish student learning and the value of teaching. A good
education system benefits society while an inferior education system hurts the community.

**Research Problem and Research Question**

There is a problem in academia. Some professors are reducing the rigor of courses in
order to influence students and receive positive student evaluations (Rice, 1988). This problem
has negatively impacted students because this reduces their actual learning and the value of a
degree.

The purpose of the research is the investigation whether some faculty change their
teaching to make courses easier and gain favor with their students with the goal of receiving
positive student ratings. The study is a descriptive phenomenology conducted at a college in
Michigan. Study participants were interviewed about their actions relating to student evaluations.

This study focused on one research question:

1. In what ways and to what extent might faculty compromise good pedagogical
   practices for positive student evaluations of teaching?

The research question examines the consequences and importance of student evaluations
of faculty. The research question asks how faculty react to student evaluations. Colleges and
universities use student evaluations as the primary tool to assess faculty performance (Clayson &
Haley, 2011).

The research question speaks to the intellectual goals of determining if higher education
is treating students like customers, ascertaining if evaluations contribute to grade inflation, and
analyzing if the process of evaluations is reliable and valid. Student evaluations affect faculty, students, administration and the general public. The purported link between evaluations and pedagogy requires investigation.

**Definition of Key Terminology**

**Descriptive Phenomenology** - is an inductive qualitative research methodology with its foundation in the philosophical tradition of Edmund Husserl. Husserl founded phenomenology and he believed that phenomenology suspended all beliefs, was linked to consciousness, and was based on the meaning of the individual’s experience (Reiners, 2012). Study participants give rich accounts of their lived experiences and researchers attempt to reveal the meaning of these phenomena while setting aside or bracketing any prior understanding. Phenomenology is descriptive when it describes an experience instead of interpreting it.

**Cognitive Consistency Theory** – a theory that suggests a person always seeks consistency in life where their behaviors and beliefs are congruent.

**Theoretical Framework**

Cognitive consistency theory (Festinger, 1957) was chosen as the framework to examine professor behaviors related to student assessments. Cognitive consistency theory states that people want equilibrium and consistency in their lives and they will change their behaviors to attain it.

Cognitive consistency theory evolved from cognitive dissonance theory of Festinger (1957) and Heider (1958). Cognitive dissonance theory centers on the consequences of positive and negative outcomes to reduce dissonance while cognitive consistency theory is slightly different. It focuses on the cognitive balance that individuals seek when inconsistencies cause tensions and force their brains and bodies to react. Behavior that differs from an established
attitude creates a need for change. This involves changing the attitude to align with the behavior. It could result in changing the behavior.

Cognitive consistency theory states that man wants equilibrium in his life (Festinger, 1957). Something happens in a person’s life that creates conflict—a inconsistency between behaviors and beliefs. This inconsistency will motivate a person to act. Man will then change his behaviors or beliefs to attain equilibrium.

There are three steps in cognitive consistency theory (Festinger, 1957):

1. People expect consistency in their lives
2. Inconsistencies create conflict or dissonance
3. This conflict or dissonance drives the person to restore consistency.

People want stability in their lives. They have expectations about people and relationships, their workplace and the world in general. A disruption or crisis in one’s life makes people want to return to constancy or equilibrium. Awa and Nwuche (2010) proffer that “Man loves the familiar; the unfamiliar is always discomforting and disturbing” (p. 44).

Cognitive consistency is a way to explain the behavior of certain faculty in relation to student evaluations. Equilibrium in a professor’s life is represented by their positive feelings about subject knowledge and teaching ability. A good student evaluation acts as confirmation and external evidence of a professor’s teaching and subject expertise.

Poor student evaluations submit that a professor is not doing a good job of teaching. This is contrary to the professor’s cognition. This causes the professor to feel negative and experience cognitive dissonance. Dissonance causes people to restore consistency. According to cognitive consistency theory, professors are motivated to change behaviors to eliminate any dissonance. These behaviors are represented by their pedagogy. Some professors may change their teaching
to gain favor with the students and get a positive rating – returning them to cognitive consistency. This compromise of a faculty member’s instruction represents a negative result for students and academia.

**Critics of Cognitive Consistency Theory**

A number of researchers have challenged the work of Festinger and other consistency theorists (Bem, 1967; Aronson, 1969; Cooper & Fazio, 1984). Bem (1967) postulated that subjects never consider their attitudes or whether they were in cognitive conflict. He suggested that a subject’s attitude was a result of a previous action and not based on conflict or dissonance resolution.

Aronson (1969) refined the theory of cognitive dissonance. He believed that dissonance caused psychological discomfort. Appearance is important to man and even though man may not be rational all the time he wants to appear rational. One is then motivated to change the behavior or the attitude for a positive persona.

Cooper and Fazio (1984) reviewed Festinger’s dissonance theory. Festinger (1957) had originally conceptualized dissonance as a motivator for a change in cognitions. Cooper and Fazio (1984) felt that both dissonance arousal and dissonance motivation were necessary for a person to act. They believed that the arousal stage was initiated by some anticipated negative consequence and that motivation was “the pressure to change one’s attitudes” (Cooper and Fazio, 1984, pg. 257). The scientists believed that dissonance arousal was important and necessary for the dissonance process but it is the psychological discomfort that precipitates an attitude change. Cooper and Fazio (1984) developed a four-step process regarding dissonance and attitude change:

1. The individual must be aware that their attitude could lead to negative consequences
2. The individual must take accountability for their action
3. The individual must feel psychological arousal
4. The individual must attribute the arousal to the action

Cooper and Fazio (1984) believed that all four steps had to be present for dissonance to occur. They opposed the concept of cognitive consistency and its tenet of change to reduce dissonance. Each researcher believed that man was motivated to act by potential negative consequences, a threat to self-image, or feelings of psychological discomfort (Cooper & Fazio, 1984).

Rationale

Cognitive consistency theory is applicable to teacher evaluations. A professor gets a poor evaluation from his students. The professor cannot accept a poor assessment. This creates cognitive dissonance and conflict. He could ignore the rating which isn’t probable or he could change his behavior. Cognitive consistency theory suggests that he would change his pedagogy in order to get a good student rating. Huemer (1998) and Germain & Scandura (2005) submit that the faculty member would dumb down the course, make tests easier, or demand very little from students to gain a good teaching evaluation. Students will rate him positively now because they do better in his class.

Professors soon recognize that by reducing the rigor everyone becomes happier. Students are gratified and contented students make for a joyful work environment. Administrators enjoy better graduation rates and more student retention. Faculty are better-off because they now earn tenure and salary increases.

Leon Festinger’s theory of cognitive consistency best informs the research question for this study. The theory states that behavior that is at odds with the established attitude demands
change (Heider, 1958). A professor’s internalization about a poor evaluation will conflict with his belief that he is a good teacher. There will be a change in behavior to ease the tension. A reduction in academic challenge hurts the students by reducing the actual learning that takes place in the classroom.

Student evaluations of faculty performance are important. A positive evaluation may lead to a reward such as a salary increase or receiving tenure. A negative evaluation may result in some sort of punishment from administration such as dismissal or a decision to not offer future employment or tenure. Faculty understand these potential consequences so there is an emotional reaction to assessment results. Cognitive consistency theory best addresses faculty behavior as a consequence of student evaluations. This theoretical framework is best suited to examine the proposed problem of practice because it explains why faculty behavior changes. A faculty member could use the formative results to improve his teaching. These actions would be beneficial to the academic institution. A professor could inflate grades or make the course easier to curry favor with the students. This reaction would be problematic for the school.

**Application of Theory to the Study**

Cognitive consistency theory strengthens the study because it explains why faculty behave a certain way after getting a negative student evaluation. Student evaluations prompt certain actions and these behaviors can be analyzed and explained within the context of cognitive consistency theory.

Cognitive consistency theory also conceptualizes the problem of practice. The tenets of this theory explain professor behaviors before and after student evaluations. Faculty behavior can also change before a student appraisal and not just after one. Faculty could alter behaviors in
anticipation of a student evaluation in order to influence the results of that appraisal and cognitive consistency theory can be used to explain this aspect of the study.

Cognitive consistency theory serves as a lens to examine the research question “In what ways and to what extent might faculty compromise good pedagogical practices for positive student evaluations of teaching?” A student evaluation may lead to some action by a professor if there is some inconsistency between a professor’s beliefs and perceptions and the results of the evaluation. A faculty member will review his previously held beliefs about himself and act in some way to eliminate the inconsistency. Cognitive consistency theory was chosen for this investigation because it offers a credible explanation for some faculty behaviors concerning student assessments.
Chapter 2: Review of Related Literature

The existing literature about student evaluations of faculty and how they impact professor pedagogy was reviewed. Student evaluations of teaching are the most common means of measuring professor teaching effectiveness (Spencer & Flyr, 1992; Seldin, 1984). A thorough investigation of existing literature was necessary to understand the impact of student evaluations on faculty teaching.

Most colleges and universities utilize student evaluations of faculty (Huemer, 1998). Almost one hundred percent of business schools in the United States use student evaluations (Clayson & Haley, 2011). Student evaluations help administrators know if faculty are competent as educators (Fink, 2008). Some researchers suggest that student evaluations are problematic and they contribute to grade inflation and a reduction in real learning (Germain & Scandura, 2005). Other investigators have offered that students are capable as evaluators of faculty teaching (Wachtel, 1998; Ory & Ryan, 2001).

The connection between student evaluations and supposed grade inflation is significant. Students that obtain a degree without the requisite skills and knowledge harm education, future employers, and society. Grade inflation decreases the value of education and harms a nation that depends on a vibrant academia.

The problem of practice is that some professors may be compromising their pedagogy by decreasing the rigor of their courses in order to influence student evaluations (Goldman, 1985). This decreases the real learning that takes place. This study adds to the existing knowledge about faculty behavior as a consequence of student evaluations.
An abundance of research on the usefulness of faculty evaluations exists. The literature review included an examination of the present research on the design, process, validity, and reliability of faculty evaluation forms. This review included studies that support the use of student assessments in the classroom and research that suggests student evaluations of faculty are detrimental to academia.

**Evaluation Form Design, Process, Reliability and Validity**

**Evaluation Form Design**

The design of professor evaluation forms is important because they are used in quality assurance procedures (Kember & Leung, 2008). They are utilized to determine whether teaching and learning meet an acceptable standard. A goal of evaluation form design should be to create an instrument that measures teaching performance and excludes irrelevant information. This can be accomplished by creating pertinent questions and having a well-designed form. The literature did not reveal an ideal format for student evaluations. Many researchers like Algozine et al (2004) discovered some common characteristics of faculty evaluation forms when researching rating instruments. They found that most forms have closed-ended and open-ended questions about faculty classroom performance and course substance. They also established that at least one question addresses the overall effectiveness of the course. Algozine et al (2004) also found that most evaluations request some written remarks about the professor and the course. They also found that anonymity for the respondent was promised by the institution and it was determined that most surveys were conducted at the end of the semester. The professor was usually absent from the classroom during the assessment.

Cashin (1995) found that student evaluations usually measure six common areas:
1. Course content
2. Workload and course difficulty
3. Professor’s ability to communicate
4. Professor-student interaction
5. Assessment details
6. Student learning

Gravestock and Gregor-Greenleaf (2008) surveyed 22 colleges and universities in the United States and Canada and found that the accountability for form design/approval varies widely. They discovered that it can range from individual faculty at Amherst College, Massachusetts to the Senate/Academic Council at the University of British Columbia. Other institutions like Harvard and University of Michigan have mandated common forms across their institutions. Evaluation form design is important since the wording, layout, and scales used can influence the results (Gravestock & Gregor-Greenleaf, 2008). There are a myriad of faculty rating forms in higher education and they can be divided into two basic types by their origin. Many forms are developed in-house while others are professionally designed and subject to psychometric testing (Ali & Sell, 1998).

An examination of the literature reveals that in-house evaluation questions can be created by the institution or sourced from existing forms of other colleges and universities. This is no guarantee of quality. Ory and Ryan (2001) are critical of institutions that don’t devote the appropriate resources to evaluation form content and design. They suggest that “many of the forms used today have been developed from other existing forms without much thought to theory or construct domains” (page 32). Marsh and Roche (1997) reported on the critical nature of evaluation form design. They suggest that the choice of questions is important to make sure that
faculty evaluations truly measure teaching performance. They posited that “Poorly worded or inappropriate items will not provide useful information” (page 1187). Marsh and Roche (1997) also examined the kinds of questions in an evaluation. They reviewed the benefit of global questions versus asking multiple questions regarding faculty classroom performance. Their conclusion was that global questions did not uncover enough detail about the many different aspects of university teaching. Other researchers like Cashin and Downey (1992) believe that short global faculty evaluations can produce good summative information. They held that a few summative questions could be an effective means to determine if a student had problems with the professor.

Evaluation forms usually have both quantitative and qualitative measurements. The quantitative component uses rating scales where the faculty member is rated according to some numerical value. Examples of the most popular quantitative measures are: Behaviorally anchored rating system (BARS), LIKERT scale, and behavioral observation system (BOS). A BARS scale has a description of a specific behavior linked to each point in the scale. A LIKERT scale is similar to a BARS scale where the student is requested to value a classroom behavior where 1= poor and 5= excellent. A BOS scale is a sliding scale where students rate how often professors exhibit certain classroom behaviors. A 1 on the BOS scale may mean that a professor exhibits a certain behavior 0% of the time and it ranges to a 5 where the instructor acts a particular way 100% of the time. A qualitative measurement is usually represented by some open-ended questions that require more than a yes or no answer. It could also be represented by a response from the student where 1 = strongly disagree and ranges to a 5=strongly agree.

Researchers do not agree on the benefits of BARS versus Likert scales. Eley and Stecher (1997) reviewed three studies where BARS was found to be less reliable and more prone to error
than Likert scales (Bernardin, Alvares, & Cranny, 1976; Reardon & Waters, 1979; Kinicki, Bannister, Hom, & Denisi, 1985). Eley and Stecher (1997) suggested that BOS was superior to Likert scales because the response scale was more complete. Academia cannot agree about assessment form design. A multitude of evaluation form designs has led to a diversity of opinion about their value and utility.

A gap exists in the current literature regarding evaluation questions because there is a lack of a universal definition for “good teaching.” Anderson, Choi, and Hair (1975) suggest that many stereotypes exist about the definition of effective teaching. They suggest that students are unsure of what they are measuring. It is difficult to assess good teaching when no one can adequately define it (Fink, 2008; Kulik, 2001).

**Evaluation Process**

The process is similar in most post-secondary institutions. It is customary for a faculty member or his designate to distribute paper-based questionnaires to the students. The surveys are usually conducted face-to-face in a classroom and the students input their answers on paper forms. The survey responses are anonymous. The surveys are manually summarized by Human Resources or an academic department (Layne, DeCristoforo, & McGinty, 1999).

Paper survey forms can be expensive when you consider the cost of paper, printing, distribution, collection, scanning, and record storage (Hmieske & Champagne, 2000). A criticism of paper-based evaluations is that professors can manipulate ratings by commenting during the distribution of forms (Simpson & Siguaw, 2000). Some instructors may modify the ratings before they return the questionnaires (Ory, 1990). In-class surveys may also be time constrained which prevents students from giving complete answers or additional comments (Layne, DeCristoforo & McGinty, 1999).
This traditional method is complemented by computer-mediated communication (Barkhi and Williams, 2010) where a student responds to an e-mail requesting an evaluation of a faculty member. This method of on-line evaluation is gaining in popularity (Barkhi & Williams, 2010). On-line evaluations can offer some advantages in terms of cost. The administrative costs of conducting paper-based surveys can be avoided by using the computer. On-line evaluations do not usually impose time limits (Layne, DeCristoforo & McGinty, 1999) and as a consequence students can provide extra comments to open-ended questions (Ha & Marsh, 1998). Barkhi and Williams (2010) reported that on-line evaluations are immune from recency bias and from the physical presence of the instructors. Sproull and Kiesler (1991) suggested that on-line evaluations may encourage a reduced sense of responsibility and students may become very aggressive when it comes to negative comments about an instructor. Sproull and Kiesler also posit that objectivity is diminished and students may respond more extremely in an on-line environment. Barkhi and Williams (2010) concluded in a study that face-to-face ratings were higher than on-line evaluations. They concluded that the lack of social presence during on-line ratings resulted in harsher assessments.

The process can be affected by a number of actions by a professor wishing to influence the student ratings. A professor can use recency bias to manipulate survey results. Students remember what was said lately as opposed to what was discussed at an earlier juncture of the course (Murdock, 1962). The faculty member can announce that he is dropping the low test score just prior to distributing the student evaluations. That strategy will elicit higher student ratings (Barkhi & Williams, 2010). A professor can also influence student evaluations by remaining in the classroom or in close proximity during the evaluations (Dommeyer, Baum, Chapman, & Hanna, 2004). Darby (2006) reported that students may try to act in a socially responsible way
by providing positive faculty ratings when the professor is close. A number of variables can affect the student ratings such as timing, layout, question design, and response rates (Ku, 2003; Ha & Marsh, 1998).

The results and the process can be influenced by the timing of student evaluations. Faculty evaluations can take place at the end of the semester or concurrently. McNulty, Gruener, Chandrasekhar, Espiritu, Hoyt, and Esninger (2010) found that the frequency and quality of student comments increased when evaluations were conducted during the semester. Early semester evaluations offer some advantages to the professor and the students (Keutzer, 1993). The professor can take the results and make necessary improvements during the semester. The students will feel some autonomy when it comes to course design. It allows the professor to make current specific behavior changes rather than wait until the end of the semester when it is too late. The evaluations go directly to the instructor which allows for quick action (Keutzer, 1993). A review of the literature suggests unanimity about the positives of early evaluation (Cohen, 1981; McGowan & Osguthorp, 2011; Lewis, 2001). Existing studies do not explain why so few universities utilize early faculty assessments. End of course evaluations are used more often than early course or concurrent assessments.

Reliability

Phelan and Wren (2005) define reliability as “The degree to which an assessment tool produces stable and consistent results” (pg.1). Others offer that reliability is enhanced by the use of an evaluation form that is well-designed and psychometrically tested (Aleamoni, 1987; Marsh, 1987; Centra, 1993). Reliability can include inter-rater reliability, test-retest reliability, and internal consistency. Inter-rater reliability is agreement or consensus of student ratings about a professor. It is an important measure of the reliability of student evaluation forms (Marsh &
Studies by Marsh (1987) suggest that the inter-rater agreement was high when he examined several examples of the average class response. He also reported that reliability is reduced when there are fewer students in a class. Marsh (1987) had research results with correlations of .95 for a class of 50 students. The correlation dropped to .60 for classes with only 5 students. Costin, Greenough and Menges (1971) and Aleamoni (1987) found the average reliability of professionally designed forms to be approximately 0.90. Test-retest reliability is gauged by giving an identical test over a period of time to the same students (Phelan & Wren, 2005). The results are compared to determine the test stability (Phelan & Wren, 2005). It may be difficult or too expensive to gather additional student evaluations about the same professor in the same semester.

Internal consistency is best measured using Cronbach’s alpha coefficient (Cronbach, 1951). It can be used to estimate the reliability of a set of questions created to measure a particular construct. This calculation determines the internal consistency of a psychometric test. A high alpha is an indication of high internal consistency while a low alpha means poor or unacceptable internal consistency (Miller, 1995). Kember and Leung (2008) suggest that adding more questions to an evaluation tends to artificially increase alpha while a sample with just a few questions can decrease it.

There have been multiple studies that suggest there are reliability issues with faculty evaluations (Marsh 1987; Burdsal & Bardo 1986; Ting 2000). Zhao and Gallant (2012) examined these findings and found they were attached to a single evaluation form and were not representative of the majority of assessment forms in the field. Earlier studies support the reliability of faculty evaluation forms (Barnes & Barnes, 1993; Feldman, 1989). These results are supported by more current studies that find that faculty evaluation forms are reliable and valid.
measures of teaching (Abrami, 2001; Wachtel, 1998; Theall & Franklin, 2001). Reliability is required in order to have validity.

**Validity**

Ali and Sell (1998) define validity as “the degree to which a test actually measures what it is supposed to measure” (pg. 4). There are a number of measures of validity. One is face validity where the questions on an assessment refer to the facet of teaching being measured (Kember & Leung, 2008). Another is content validity. Content validity is best measured by subject experts that can offer an opinion whether survey questions cover the required content. Marsh (1987) suggests that content validity represents the best avenue for validation of student evaluation forms (Kember & Leung, 2008). Kulik (2001) and Abrami & Mizener (1983) purport that the validation of student evaluation forms is problematic because there is no single definition for effective teaching.

Many researchers have deemed student evaluation forms to be valid (Marsh, 1987; Feldman, 1989; Cohen, 1981). The validity of student evaluations has been supported through correlations with measures of student learning, peer ratings, alumni ratings, and trained observers’ ratings (Arreola, 1995). d’Apollonia and Abrami (1997) researched the validity of student ratings by examining student learning. They looked at whether there was a correlation between higher-rated professors and higher student grades. Their study was based on a multi-section validity design that compared the results of multiple sections of a course taught by different professors using a common exam (d’Apollonia & Abrami, 1997). Their conclusion was that there “was a moderate to large association between student ratings and student learning” (1997, pg. 1202). They also concluded that student evaluations of instructors were valid.
indicators of student learning (1997, pg. 1202). Their conclusions were supported by similar studies and findings (Cohen, 1981).

Scriven (1983) does not support the notion of the link between student learning and student evaluations. He contends “the best teaching is not that which produces the most learning” (pg. 248). He suggested that good marks may be dependent on factors not related to good teaching. He gives an example of a professor that pressures his students until they effectively abandon other courses, leaving more time for them to study for his class. Scriven (1983) points out in this case that good test scores can result from unethical professor behavior. Arreola (1995) and Costin, Greenough, and Menges (1971) suggest that student ratings of instruction are comparable to colleague ratings. Arreola (1995) reported that student ratings had correlations of 0.87 to 0.89 from semester to semester while Costin, Greenough and Menges (1971) had correlations of 0.70 to 0.87 for the same professors and courses.

Murray (1983) undertook a study where he trained observers to rate certain teaching behaviors of 54 professors. These instructors had been rated from low to high performers by their students. Each instructor was observed for a total of 18 to 24 hours during the semester. Murray (1983) concluded that the trained observers and students had similar assessments of faculty classroom performance. Feldman (1989) reviewed Murray’s study and an additional four studies that correlated ratings made by observers and student ratings. He found the average correlation coefficient between observer and student assessments to be .50. His study also suggests that that observer and student ratings correlate well.

The validity of student ratings has been upheld when alumni assessments of faculty are examined. Overall and Marsh (1980) conducted a longitudinal study about alumni evaluations. All of the students completed assessment forms for each course they took during a three year
period. One year after graduation they were again requested to complete evaluations. The original ratings in one hundred courses correlated .83 with the alumni assessments (Overall & Marsh, 1980). Kulik (2001) suggested from his research about the validity of student ratings that “current students give favorable ratings to teachers whom alumni remember fondly and poor marks to teachers whom alumni remember unfavorably” (pg. 4). Kulik (2001) also concluded from his reviews on the validity of student ratings that some alumni have given interviews and wrote about former teachers and these comments were consistent with current student ratings. Valid evaluations are dependent on the assessment forms measuring what they are supposed to measure: teaching effectiveness. Marsh (1987), Murray (1984), and McKeachie (1979) have concluded that student evaluations are valid. Greenwald and Gilmore (1997), Ambady and Rosenthal (1992) and Williams and Ceci (1997) suggest that ratings do not measure teaching effectiveness but assess grading habits and teacher personality.

Academia needs a definition for effective teaching to facilitate the student measurement of faculty performance. Teaching has many components like rigor, learning climate, subject expertise, and student engagement. More study is warranted to determine what constitutes effective instruction.

Conclusion

A synthesis of the literature supports the notion that evaluations forms are generally well-designed and measure faculty classroom performance. Studies submit that the process of evaluating faculty is fair and similar in most schools, and the forms are usually reliable (Theall & Franklin, 2001) and valid (d’Apollonia & Abrami, 1997). A fewer number of studies had different conclusions about the form design, process, reliability, and validity (Ory & Ryan, 2001; Simpson & Siguaw, 2000; Arubari, 1987; Greenwald & Gilmore, 1997). This smaller group of
researchers postulated that many evaluation forms were of poor design, the process was flawed and subject to manipulation by faculty, and the ratings were not reliable or valid. The literature advocates that any problems linked to faculty evaluations do not have their origin in form design or process.

**Support for the Use of Professor Evaluations**

Student evaluations of teaching began in earnest in the late 1960s and early 1970s (Murray, 2005). This initiative was supported by students because they wanted input into how they were taught. Evaluations were mandated by administrators concerned with accountability and a desire to improve professor performance (Murray, 2005). The purpose of faculty evaluations was initially to assist professors to improve their classroom instruction and to help measure the quality of faculty performance (Berk, 2005). Tighter education budgets and a public demand for value for their tax dollars have now made teaching quality a priority (Frost & Fukami, 1997). A consequence of this public scrutiny is that faculty evaluations are deemed necessary (Foote, 1998).

Student evaluations of faculty are used in the majority of universities and colleges (Kelly, 2012; Huemer, 1998). Most schools today use student evaluations of faculty as their primary means of rating professors (d’Apolollonia & Abrami, 1997). Loeher (2006) surveyed 62 Association of American Universities and found that “98% of them used student questionnaires as the primary method of evaluating teaching” (p.2). Nearly one hundred percent of business schools in the U.S. mandate the use of faculty evaluations (Clayson & Haley, 2011). Studies suggest that students are the most important source for rating faculty (Scriven, 1995; Rifkin, 1995). Stocham and Amann (1994) suggest that using students to evaluate faculty signifies a belief by their institution in the following:
1. Student involvement is fundamental to active learning.

2. Perceptions of faculty are influenced by student personality and behavior.

3. Students are active learners which provide opportunities for faculty growth and change.

4. Students are important in the teaching-learning environment.

5. The teaching-learning dynamic is constantly changing.

   Scriven (1995) suggests that students benefit from their unique position in academia.

   They are best to make judgments about:

   1. Their increased understanding about a subject
   2. Their change in motivation about a course, related careers, and further learning
   3. Professor behavior like absenteeism or punctuality
   4. Faculty teaching styles
   5. Test composition and grading
   6. Relevant information unrelated to teaching like textbook expense

   Literature as far back as the 1970s (Goldschmid, 1978) suggests that students are competent as evaluators of faculty performance. Many recent studies support that notion (Theall & Franklin, 2001; Wachtel, 1998; Nasser & Fresko, 2002; Ory & Ryan, 2001). Scriven (1995) posits that students can assess whether or not their knowledge and comprehension has increased which is directly related to the teaching ability of the professor. Murray (2005) suggests that students can evaluate classroom performance but cannot assess non-classroom behaviors like course design, professor knowledge, and publishing. Some research contends that student evaluations are incomplete and must be complemented by other means like peer, administrator, alumni, or trained observer assessment (Murray, 2005; Scriven, 1995).
Formative Results

Research studies account for a number of positive benefits from the use of student evaluations of teaching. One benefit is the development or self-improvement of professors (Centra, 1994; Marsh & Dunkin, 1997; Stapleton & Murkison, 2001; Simpson & Sigquaw, 2000). Faculty can take the formative results from teaching assessments and change their classroom performance or they can strengthen good practices (Centra, 1993). Faculty will benefit from student evaluations if:

1. They learn new information from the results.
2. They appreciate the results.
3. They want to improve.
4. They have the ability to change their classroom behaviors.

Most universities conduct student evaluations at the end of the semester (McNulty et al, 2010). Some researchers suggest that faculty can benefit even more if they are given the formative results from early semester evaluations (Keutzer, 1993; McNulty et al, 2010; McGowan & Osguthorp, 2011; Lewis, 2001). This gives the professor an opportunity to change classroom behaviors and improve teaching during the semester rather than waiting until the end. This will directly impact student learning. It also gives the students a chance to design their own instruction. Assessments made early in the semester are more useful to faculty compared to a global assessment at the end of the term (Keutzer, 1993).

Summative Results

Student assessments have increased in significance because they provide impartial data for faculty appraisals (Williams & Ceci, 1997). Faculty evaluations meet the requirement for
greater accountability in teaching. Administrators find student evaluations convenient because they are impartial, easy to use and inexpensive (Seldin, 1984). Student evaluations also provide numerical scores that are objective and reliable (Williams & Ceci, 1997; Stapleton & Murkison, 2001). Student assessments also transfer some of the responsibility for judging professors from the administrators to the students (Johnson & Kelley, 1998).

Administrators can use the summative results to make decisions about hiring, promotions, pay raises, teaching awards, tenure, and professional development (Beran & Violato, 2005; Arreola, 1995). Beran et al. (2007) discovered in their study that 44 out of 52 university administrators used evaluation results for personnel decisions. Some institutions like Ryerson and Harvard use the written comments on faculty evaluations to nominate professors for teaching awards and for future reference for promotion or tenure (Gravestock & Gregor-Greenleaf, 2008). Some schools publish the results and students can use the information to select courses (Beran, Violato, Kline, & Frideres, 2009).

Conclusion

Student evaluations of teaching have been popular since the 1960s (Murray, 2005) and they were originally designed to improve accountability and faculty performance. Student assessments abound and they are the primary method to evaluate teaching (Loeher, 2006). They provide summative results that administrators can use to make human resource decisions like promotions or raises ((Williams & Ceci, 1997). Evaluations also provide formative information that a professor can use to improve his teaching (Simpson & Sigquaw, 2000).
A literature review revealed that there have been over 2000 studies on student evaluations since the 1920s (Wilson, 1998). A synthesis of the literature discloses that a minority of these reports support student evaluations.

**Opposition to Student Assessments**

A literature review on this topic suggests that many researchers have found fault with student evaluations of faculty. Their opposition is based on a number of criteria like the reliability and validity of the forms, the process, and the inability of students to be effective raters.

**Reliability and Validity**

Some research about the validity of rating forms has reached different conclusions. Greenwald (1997) found that the validity of evaluation forms was disputed in the 1970s but by the 1980s “most expert opinion viewed student rating measures as valid and as worthy of widespread use” (pg. 1182). Ali and Sell (1998) found that some studies have suggested that student evaluations of teaching are valid and reliable (Murray, 2005; Greenwald, 1997; Aleamoni, 1987; Arreola, 1995).

Balam and Shannon (2010) discovered that class size, gender, appearance, personality of the professor and chosen academic major were threats to the validity of student evaluations. They found that students constantly had a bias when it came to assessing instructors based on these factors and they suggested that teaching effectiveness was not being measured accurately. A literature review on studies critical of student evaluations provides some context to the current debate. The following sections represent the main areas of concern:
Gradning Leniency Bias

Researchers suggest that students give better faculty reviews when they expect to get a higher grade in the course (Rice, 1988; Cahn, 1987; Shapiro, 1990; Benz & Blatt, 1996). Huemer (1998) states that “this correlation is well established” (pg. 3) and posits that student evaluations reflect grading leniency as much as teaching skill. Huemer (1998) also offered that good marks from a professor was the same as getting praise and people like those that flatter them. Marsh (1987) researched student evaluations and found that a majority of professors believed grading leniency resulted in better student evaluations.

Grade inflation is an acknowledged problem in higher education (Winship, 2011). Student grades have been on the rise since the 1960s (Astin, 1998). Studies have linked grade inflation to student evaluations (Stumpf & Freedman, 1979; Greenwald, 1997). Pfeffer and Fong (2002) reported that “Grade inflation is pervasive in American higher education (pg. 83). Babcock and Marks (2010) did an investigation on how much time students devoted to weekly study. They found that over the last 40 years study time had declined by 50%. They also found 35% of students surveyed reported a weekly study time of 5 hours or less. They concluded that students did not have to study because of grade inflation. They received good grades with little effort and they were not motivated to do more.

Dumbing Down Courses

Related to grading leniency and grade inflation is the dumbing down of courses. Students achieve higher marks when a course is less challenging and the professor subsequently gets higher student ratings. Ryan, Anderson, and Birchler (1980) conducted a survey and 38% of responding faculty admitted to making courses easier in order to get higher student ratings.
There is pressure on faculty to make students “happy” since this is a focus of many colleges and universities. Winship (2011) and Abrami, d’Appolonia and Cohen (1990) argue that student evaluations do not measure teaching performance but attempt to determine student satisfaction. They also maintain that faculty feel pressure to get better evaluations to earn rewards like pay raises and tenure. Faculty can reduce the rigor which results in higher marks and better evaluations.

**Student as Customer Metaphor**

Huemer (1998) and Abrami et al (1990) suggest that student evaluations of faculty are effective for measuring student happiness. The idea that post-secondary schools are businesses and their focus should be the customer is commonplace on campuses today (Germain & Scandura, 2005). Baldwin (1994) argues that “teaching as an activity which is conducted for profit will result in bad teaching” (pg. 136). McMillan and Cheney (1996) opine that the student as customer metaphor has developed into a negative. They suggest that students often exhibit little interest in coursework and they expect to be entertained. They submit that students expect high grades with little work. Researchers comment that unhappy students complain about tuition fees like they would complain about any major purchase (McMillan and Cheney, 1996; Cote and Allahar, 2007; Newson, 2004). Huemer (1998) agrees that student satisfaction should be a goal of the university but it must not be more important than educating the students.

**Rating Bias**

A review of the literature about rating bias suggests it is a problem with student evaluations of teaching (Huemer, 1998; Griffin, 2001; Morris, Gorham, Cohen, & Huffman, 1996; Naftulin, Ware, & Donnelly, 1973; Sherman & Blackburn, 1975). Student ratings are
affected by appearance and teaching style (Huemer, 1998). Gender plays a role in student evaluations. Female faculty are assessed lower by male students than female students (Kozub, 2010). Minority faculty tend to be rated higher by minority students (Germain & Scandura, 2005). Age has an effect on assessment results. Younger students tend to rate younger faculty higher while more mature students assess older faculty more favorably (Germain & Scandura, 2005). Other studies have suggested that student evaluations are a personality or popularity contest (Shevlin, Banyard, Davies & Griffiths, 2000; Worthington, 2002).

**Academic Freedom**

Some researchers posit that teaching evaluations lead to a loss of academic freedom for professors (Haskell, 1997). Huemer (1998) suggests that student evaluations may influence what a professor says in class. He contends that professors, in an effort to not offend anyone and get a good evaluation, may act like a politician and never take a controversial position on an issue. Dershowitz (1992) reported that students take exception to some of his comments and opinions in class. These students have threatened retaliation via the student evaluation. Dershowitz is a noted senior professor and has little fear of losing his job. That is not the case with junior faculty. Huemer (1998) suggests that faculty seeking tenure will never speak their mind on a controversial issue which represents a loss of academic freedom.

Haskell (1997) reports evidence of complaints by faculty in his study about student evaluations of faculty. Some professors are critical of evaluations and feel that they are infringements of academic freedom (Stone, 1995; Young, 1993; Haskell, 1997). These instructors say that student evaluations force conformity to political standards (Young, 1993).
Damron (1996) feels evaluations can be used to control behavior since the results determine raises or promotions.

**Conclusion**

There have been many studies about student evaluations of faculty since the 1920s (Wilson, 1998). A plethora of research exists that finds fault with student ratings of professors. (Huemer, 1998; Balam & Shannon, 2010; Marsh, 1987). A synthesis of literature suggests that student evaluations are inaccurate and the summative and formative results are suspect. The literature also revealed that student assessments are popular and remain the primary method to assess faculty classroom performance (Clayson & Haley, 2011).

**Summary**

A review of the literature revealed different opinions about nearly every aspect of student evaluations of faculty teaching. Some researchers and faculty supported the use of student evaluations. The majority of researchers did not support student assessments.

The literature did not reveal a universal form design. Some assessments are designed in-house while others are professionally crafted (Ali & Sell, 1998). Some universities and colleges have early or concurrent evaluations while the majority of schools have end-of-semester evaluations. Early studies generally support the reliability of evaluation forms (Barnes & Barnes, 1993; Feldman, 1989). They have been upheld by more current research (Abrami, 2001; Theall & Franklin, 2001).

Some researchers suggested that student evaluations are valid (McKeachie, 1979; Marsh, 1987; Murray, 1984) while others (Greenwald and Gilmore, 1997; Williams and Ceci, 1997;
Ambady and Rosenthal, 1992) submit that evaluations are not credible because ratings do not truly measure teaching effectiveness. The results of the literature review point out a lack of clarity and a gap in the research about student evaluations.

The thesis statement for this study is “Student evaluations of faculty are problematic and their summative and formative results may contain inaccuracies.” This assumption was reached after an examination of the existing literature.
Chapter 3: Research Design

The goal of this inquiry was to determine the effect of student evaluations on faculty pedagogy. The research question used in this study was “In what ways and to what extent might faculty compromise good pedagogical practices for positive student evaluations of teaching?” The research question helps to determine how some faculty react to student evaluations and how their behaviors may affect academia.

Qualitative Research Approach

A qualitative research design was chosen because it leads to an understanding of phenomenon like student evaluations. Qualitative research seeks to explore a phenomenon and to search deeper into the nuances of the problem at hand from the perspective of the participants. The account of each person’s experience is often described as “thick” (Denzin, 1989) because a researcher must understand the purpose and meaning behind the appearances and words (Moustakas, 1994). A qualitative research design links data to be collected to the original questions of the study (Yin, 1994).

Qualitative study is informed by constructivist theory founded by Jean Piaget. Constructivism is a theory of learning which emphasizes that people create meaning of their world through individual constructs. Constructivism ascribes that knowledge is built by the interaction of an individual and their social reality. Constructivists hold that reality is created in a subject’s mind (Hansen, 2004) rather than it being an outside entity. Some constructivists suggest that meaning is buried and can only be discovered by deep reflection (Schwandt, 2000) while other constructivists believe that only an interaction between the researcher and the research subject can uncover the deeper, hidden meaning of a phenomenon (Ponterotto, 2005).
Nieswiadomy (1993) maintains that the researcher must put to one side his or her own experiences in order to recognize those of the study subjects.

Qualitative research is better suited than quantitative research method in meeting certain intellectual goals of a study (Maxwell, 2005). Maxwell (2005) suggests that these goals include:

1. Understanding the significance of phenomena for the study participants
2. Understanding the background and environment for the actors
3. Identifying unexpected phenomena
4. Understanding the method by which events transpire
5. Develop contributing explanations

Qualitative research allows for the use of a theoretical lens through which all aspects of the phenomena can be examined (Miles & Huberman, 1994). This lens helps to determine the kind of questions that are asked, guides the data collection and analysis, and helps frame the research findings (Anfara & Mertz, 2006). The use of a theoretical lens helps guide researchers to the important issues and which actors should be studied.

Qualitative inquiry was appropriate for this study because of its inductive approach, its emphasis on people and its focus on dialogue. Qualitative research is best suited for understanding meaning, the participants’ environment, and discovering unanticipated phenomena and influences.

Phenomenological research method was used in this investigation. Edmund Husserl was significant in the development of phenomenological research. Husserl developed a technique of reflection that described objects with as few assumptions as possible. Husserl’s method of reflection helped to describe the phenomenon by suspending presuppositions. Husserl’s phenomenology replaces interpretation with intuition.
Phenomenology examines the independence of an individual’s lived experiences and subjective reality. Moustakas (1994) describes phenomenology as a philosophy and a study technique where a small number of subjects are engaged to develop patterns and associations of meanings. Moustakas (1994) also describes phenomenological science as a relationship between the external that is observed and the internal which is judgments, discernments, and remembrances. Moustakas believed that “all knowledge and experience are connected to phenomena “p. 44) and that a unity exists between the observer and the observed. A number of activities are designed to generate phenomenological reflection and they include line-by-line reading and existential analysis.

Descriptive phenomenology was chosen as a research methodology for this study because it focuses on the subjective experience of individuals. It is an inductive qualitative research methodology with its roots in the philosophical practice of Edmund Husserl. The purpose of descriptive phenomenology is to describe phenomena rather than interpret or explain. Descriptive phenomenology is a framework for inquiry rather than a rigid process or procedure. It is a “research methodology aimed at producing rich textual descriptions of the experiencing of selected phenomena in the life world of all individuals that are able to connect with the experience of all of us collectively” (Smith, 1997, p. 80). The object of interest is experience and it must be translated into script by descriptive account (Goble and Yin, 2014). Researchers take the rich and complex evidence that is given to them and they make assertions about its cogency.

Phenomenology focuses on the trivial aspects of a lived experience and this will shape the type of questions asked in an interview. The researcher can craft questions that move from parts of the lived experience to the whole of the experience and back again. This will increase the depth of engagement between researcher and research subject and assist with interpretation.
Participants

Participants for this study were 10 full time faculty with at least three years teaching experience. Five males and five females were chosen as participants from a pool of professors teaching at a college in southeastern Michigan. Their ages ranged from 30 years to 60 years. The subjects that they taught were not a factor in their selection. The number of faculty chosen met the standard for a phenomenological study. Qualitative research requires a small sample of people so that they can be studied in depth. Creswell (1998) suggests that a minimum of five study subjects should be interviewed for a phenomenology. Morris (1994) submits that at least six people should be questioned for a phenomenological study.

Palinkas, Green, Duan, Horwitz, Wisdom, and Hoagwood (2013) state the typical manner of choosing individuals in a qualitative study is called purposeful selection or purposeful sampling. This is a requirement when using a small sample size to ensure that the population being studied is knowledgeable and experienced with the phenomena. Patton (2002) suggests that purposeful sampling makes the best use of limited resources while Maxwell (2005) offers that purposeful sampling provides confidence that the study findings represent the average members of the population.

The site chosen was a large college in southeastern Michigan. The location was a concession to geography because it is located close to the researcher’s residence. An application to gain access to the study site for purposes of research with human subjects was filed with college administration. A letter (see appendix A) was sent by the researcher to each tenure-track professor at the college detailing the importance of the study and asking for their participation. This letter detailed the measures taken to ensure their anonymity and the interview protocol. The study subjects were given assurances in writing that they would not be harmed in any manner,
they would be treated ethically, and their identities would remain confidential. The researcher used pseudonyms for all study participants (see Table 1).

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Pseudonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty # 1</td>
<td>Betty</td>
</tr>
<tr>
<td>Faculty # 2</td>
<td>Diane</td>
</tr>
<tr>
<td>Faculty # 3</td>
<td>Eliza</td>
</tr>
<tr>
<td>Faculty # 4</td>
<td>Vicky</td>
</tr>
<tr>
<td>Faculty # 5</td>
<td>Adam</td>
</tr>
<tr>
<td>Faculty # 6</td>
<td>Janice</td>
</tr>
<tr>
<td>Faculty # 7</td>
<td>Joe</td>
</tr>
<tr>
<td>Faculty # 8</td>
<td>Kevin</td>
</tr>
<tr>
<td>Faculty # 9</td>
<td>Craig</td>
</tr>
<tr>
<td>Faculty # 10</td>
<td>Jordan</td>
</tr>
</tbody>
</table>

Table 1: Pseudonyms for Study Participants

The study subjects were assured that the researcher would adhere to and obey the code of conduct for researchers as prescribed by the American Educational Research Association. This was to guard against any impropriety and misconduct by the researcher that may reflect on the institution or the study subjects. There was token compensation for study subjects and it was detailed in writing to them. The study subjects signed an informed consent document (see appendix B) before participation in the study was allowed. Each study subject received follow up correspondence thanking them for their participation in the study.

**Procedures**

**Data Collection**

The researcher is the primary instrument in data collection and data analysis. Miles and Huberman (1994) contend that evidence of a good qualitative researcher is some awareness with
the phenomena and venue. Data was collected through semi-structured, in-depth interviews and analysis of field notes and journals. Data analysis went on concurrently as data was collected. This gave the researcher an opportunity to test emerging themes and categories against subsequent data.

The interviews were conducted on-campus in a private, confidential location or at a local library. There were two interviews. Interviews are the most effective and frequent means in a phenomenological study to prompt study subjects to reveal their personal experiences and perspectives. Initial interviews were limited to one hour and a follow-up interview followed. Interviews were directed by an interview guide (see schedule D & E) to gather some common responses and to help answer the research question.

Miles and Huberman (1994) suggest that the researcher modify the recording protocol, interview, and observation techniques with each field trip in order to improve data collection. The initial interview was audio taped and in the second interview the study subjects provided written responses. The recordings were transcribed and the study subjects had an opportunity to review and approve the interview transcripts. Interviews produce a large volume of data. This rich production of data is invaluable when the data is analyzed to uncover major issues and themes (Lester, 1999).

Field notes were taken during and after the interviews to record impressions, context, and to aid reflection about the data. Field notes contained descriptive information and facts like date and time, behaviors, setting, and actions. Field notes also contained thoughts, questions, and concerns. Good field notes must be accurate, organized, focus on the research problem, and record the observer’s thoughts and insights (Emerson, Fretz, & Shaw 2011). There should also be notations on how the researcher may influence the situation.
The researcher kept a reflective journal to record experiences, assumptions, and perspectives. The purpose of a reflective journal is to record the researcher’s bias in the study and minimize the investigator’s influence in the inquiry (Ortlipp, 2008). This reflective practice demonstrates that the researcher’s experiences will have some influence on how the inquiry is conducted and how these experiences might influence the study results (Harrison, MacGibbon, & Morton, 2001). Another benefit of a research journal is an increased transparency in the investigative process. A researcher acknowledges their values instead of bracketing them out.

Data Analysis

It is the responsibility of the researcher to provide the reader with a clear process and access to the study subjects’ experiences through effective analysis of their interviews (Grbich, 2007). Analysis in a qualitative study usually includes a large volume of data. Interviews from the ten participants generated an abundant amount of information. Grbich (2007) suggests that data analysis includes:

1. Searching for themes from the transcripts
2. Reflecting on the interviews and journalizing feelings
3. Identifying the central themes

Analysis of the transcripts began with open coding. Open coding is the categorization of the data resulting from close examination (Strauss & Corbin, 1998). Open coding looks for similarities and differences in the data. Line by line analysis was effective during this initial coding stage. This step established some common categories. This first step was reviewed and reflected upon to ensure an accurate categorization of data. Axial coding followed open coding. Axial coding is the linking of codes to each other through inductive thinking. In this step sub-categories were formed from the data and this action assisted the researcher to find deeper
meaning of the phenomenon. At this juncture the data was reassembled so that the researcher could identify relationships and connections.

**Ethical Considerations**

Data storage is of paramount importance in any research study. A researcher and any study results will lose credibility if the storage of data is compromised. A number of procedures were followed to assure that all data was safely secured and stored and the confidentiality of all involved in the study was maintained.

Miles and Huberman (1994) suggest that a researcher needs a “well-organized physical place for storage and retrieval of raw field notes, tapes, edited hard copies, memos, and so on” (p. 45). This study has a lockable storage cabinet. All hard copy data was stored there and only the researcher has the key. The research study computer files are stored on the researcher’s personal computer under a coded name and the computer is only accessible by password. The computer is also virus protected. An index of all stored material is kept in the locked storage cabinet and also on the researcher’s personal computer.

**Trustworthiness**

Trustworthiness refers to the final component of the study design and whether the findings are correct from the perspective of the researcher, study subject, and the audience (Creswell and Miller, 2000). Trustworthiness represents the study’s credibility, transferability, dependability, and confirmability.

Credibility means that the study subjects and readers believe the findings. Credibility in a study can be provided in a number of ways. Using well-established research methods adds to the integrity of the study. The questioning and data analysis must be consistent with similar studies. The researcher requested that the study subjects be open and frank during each interview.
Member checking is considered to be an important aspect of credibility. The study subjects are asked to confirm their interview transcripts concurrently or at the end of an interview. The researcher also confirmed his independence to the study subjects and suggested that all answers to the questions are acceptable. Credibility was promoted by using thick description of the phenomenon and by describing the context of the data collection. Study findings are aligned with similar studies.

Transferability means the results of the inquiry are applicable outside the boundaries of the study. It is important to give a thick description of the study phenomena and enough contextual information so that readers can recognize it and see how the findings can be applied to their situation. The researcher utilized block quotes from the participants and extensive description. This helped to determine the transferability of findings to a similar circumstance.

Marchionini & Teague (1987) suggests that transferability is enhanced when the researcher defines the boundaries of the study at the outset. They advise that the number of organizations taking part, information about the study subjects, number of participants, data collection methods, number of interviews, and time frames are important for readers to measure transferability to their settings.

Dependability in a qualitative study refers to whether the findings are reliable and could be repeated in a similar context with the same methods and study subjects. This means that the processes used in a study must be described in detail so that a new researcher could arrive at similar findings, interpretations and conclusions based on the research data. Shenton (2004) suggests that dependability is enhanced by three factors:

1. A section should be devoted to the research design which explains what was planned and undertaken.
2. A detailed description of data gathering and what was performed in the field.

3. A reflective assessment of the study that includes a measure about the process of the inquiry.

This inquiry has an audit trail with field notes, documents, and reports in order have a dependable investigation. Trustworthiness is diminished without it.

Confirmability in this study refers to the degree that the results are formed by the respondents and not researcher interest or bias. Lincoln & Guba (1985) suggest four ways to create confirmability. They are:

1. Confirmability audit
2. Audit trail
3. Triangulation
4. Reflexivity

A confirmability audit is an external audit where a different researcher examines the process and findings to see if the outcomes and conclusions are supported by the data. They help to establish the accuracy of the research. The confirmability audit provides excellent feedback for the researcher about the quality of the results and helps to determine if more data is required.

An audit trail is a paper trail. It is a review of the process, data collection, analysis and conclusions of a study. An audit trail was established and allows other investigators to confirm all aspects of the research.

People use triangulation because it is better than using a single data source. Some suggest that it can be used for corroboration of data. Triangulation is using multiple data sources
in order to increase the richness of the interpretation. This investigation used interviews, field notes, and a reflective journal to corroborate data.

Reflexivity is the examination of the context in knowledge construction and the influence of the researcher in the research process. Many researchers suggest that the background of the researcher is important and that it will have an effect on the research process. Lincoln & Guba (1985) propose that a researcher should keep a reflective journal. It can describe methodological decisions and the reasons for them. It can also define a researcher’s values and interests and how the study is affecting them. Documenting one’s beliefs and preconceptions in the study improves the trustworthiness of the research process and findings.

Potential Researcher Bias

The researcher acknowledged some notions about student evaluations. The investigator had twenty-six years in education including nineteen years of teaching experience. The researcher has evaluated faculty as a student and has been rated by students for his teaching performance. Student evaluations of faculty seem flawed for a number of reasons. Students are not trained to rate professors and evaluations appear little more than popularity contests. Evaluations may actually hurt academia.

Some strategies exist to manage researcher bias and they were followed. One is having a proper study design and methodology that will stand up to scrutiny. All initial interviews were recorded electronically and the respondents were asked to review the transcripts for accuracy. Attention was paid to the order of questions so that bias was not created. Questions were neutral and there were no leading questions. Unaided questions were requested before aided questions. General questions were asked before specific queries. The researcher did not put words in the
respondents’ mouths. The researcher was mindful of his tone of voice and body language so he did not influence dialogue. A reflective journal and field notes were kept and reviewed to decrease subjectivity and maximize objectivity.

The researcher realizes that all studies are subject to some bias. Every effort was made to have a truthful, neutral study where the conclusions depended on the evidence and study subjects and not on the inquirer’s partialities.

**Limitations**

There were several limitations identified in this investigation. The first one is the narrow scope. Only one school was chosen as a research site and a small number of faculty participated in the study. This study only involves faculty from Michigan.

The demographic breakdown is limited. The study participants were only identified as 5 male and 5 female full-time professors with a minimum of three years teaching experience. Ethnicity was not known. Ages ranged from 30 years to 60 years. Eight participants had masters degrees and two study subjects had Ph.D. degrees.

**Conclusion**

The purpose of this study is to understand the behavior of professors as a result of student evaluations. This investigation was guided by the research question “In what ways and to what extent might faculty compromise good pedagogical practices for positive student evaluations of faculty?” Faculty comportment in the classroom has a major impact on real learning and academia. This study is important to faculty, students, and administrators.

Student evaluations are used in most colleges and universities in the United States (Kelly, 2012). A plethora of studies exist about the usefulness of student evaluations (Greenwald, 1997;
Santiago & Benavides, 2009). Some researchers like Murray (2005) conclude that student evaluations improve teaching. Other researchers like Huemer (1998) believe that student evaluations are problematic and should be abolished.

This study was guided by the theory of cognitive consistency (Festinger, 1957). It suggests that people want stability in their lives. A negative student evaluation may represent a psychological tension that a faculty member seeks to eliminate. A professor’s behavior as a consequence of student evaluations may be harmful, positive, or neutral. An examination of the professor’s experience is warranted.

A review of the literature about student evaluations and the faculty reaction suggests disagreement about nearly every aspect of student assessments. Some researchers like Marsh (1987) support the use of student evaluations and conclude that they produce accurate results. Others like Ambady and Rosenthal (1992) believe that student evaluations do not measure classroom performance. Other researchers have examined the evaluation process and the trustworthiness and reliability of the evaluation instrument and reached different conclusions.

A qualitative study was undertaken because of its inductive approach, its emphasis on dialogue, and its concentration on people. The method of inquiry was descriptive phenomenology. The participants were faculty from a large-sized college in southwestern Michigan. Data was collected through semi-structured, in-depth interviews and analysis of field notes plus reflective journals. An analysis of all data took place concurrently and at the end of the study. Data was safely secured and stored and the study subjects remain anonymous. A number of actions took place to ensure trustworthiness and reliability of information.
Chapter 4: Research Findings

The purpose of this research was to explore the effect of student evaluations on faculty pedagogy using cognitive consistency as a theoretical framework. The study used qualitative inquiry and a descriptive phenomenology research design to investigate whether faculty compromised good teaching pedagogy to receive favorable student evaluations.

The institute selected was a large college in southeastern Michigan. Approximately 300 full time faculty were invited by college e-mail to participate. Ten subjects were chosen from the faculty respondents. Five male faculty and five female faculty participated in the study. Their ages ranged from 30 to 60 years old. Eight participants held masters degrees and two held Ph. D. degrees.

The participants for this study are employed under a union collective agreement. They are evaluated by their students for every course at least once per year. The student evaluation form is standardized but can be amended by the faculty member and / or their department.

The participants are full time faculty and evaluated by their department once every three years. New hires are on probation for three years and are evaluated by students for each class during every semester. Reviews include a written self-appraisal, a description of goals and strategies to achieve them, a peer review, student evaluations and a Dean classroom visit. There is no contractual requirement to review student evaluation results for full time faculty more frequently than once every three years. There is no weight attached to any of the evaluation methods.

Data was collected using in-depth, one-on-one interviews which produced rich descriptions of the study subjects’ experiences with student evaluations. Data analysis took place after consideration of the research objectives and an inductive interpretation of the raw data.
produced by readings of the interview records and an examination of the field notes. Each transcript was reviewed multiple times to determine what each faculty member felt and how they experienced it.

Two superordinate themes were developed. A multiple depiction of the essence of the phenomenon consists of two primary themes and ten corresponding supporting themes: 1. Faculty Views of Student Evaluations, including a. accuracy, b. significance, c. role, d. process, e. student training; 2. Faculty Experience with Student Evaluations, including a. opinions, b. emotions, c. preparation, d. formative, e. administrative action. The following table represents the two core themes and ten supporting themes:

<table>
<thead>
<tr>
<th>Themes</th>
<th>Subthemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty Views of</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Student Evaluations</td>
<td>Significance</td>
</tr>
<tr>
<td></td>
<td>Role</td>
</tr>
<tr>
<td></td>
<td>Process</td>
</tr>
<tr>
<td></td>
<td>Student training</td>
</tr>
<tr>
<td>Faculty Experience</td>
<td>Opinions</td>
</tr>
<tr>
<td>With Student Evaluations</td>
<td>Emotions</td>
</tr>
<tr>
<td></td>
<td>Preparation</td>
</tr>
<tr>
<td></td>
<td>Formative</td>
</tr>
<tr>
<td></td>
<td>Administrative Action</td>
</tr>
</tbody>
</table>

Table 2: Superordinate Themes and Sub-Themes

The interviewees’ accounts about the phenomenon had similarities. Their experiences could be grouped into two superordinate themes and ten subthemes. This chapter will introduce the themes and subthemes supported by narrative and direct quotes.

**Student Evaluations**

Interviews with study participants yielded a number of key themes and subthemes. One key theme was the faculty view on student evaluations. Student evaluations are the primary
means to evaluate faculty at colleges and universities. An evaluation has many facets and each one is important. Discussions with faculty about student evaluations helped to identify a number of subthemes. These subthemes included accuracy, significance, role, process, and student training.

Five male and five female full time faculty with at least three years teaching experience were chosen to participate in this study. Pseudonyms were assigned to each participant to maintain their anonymity.

Disagreements exist about the importance of student evaluations. Many professors believe that they are significant while others do not. The interviewees for this research tended to attach more importance to Dean evaluations, peer reviews and self-assessment. Some faculty used the term “useless” to describe student evaluations which suggested a history of bad experiences with student assessments.

Accuracy

Accuracy is a subtheme of the major theme faculty views on student evaluations. Eight out of ten research subjects said student evaluations were not accurate. The validity and reliability of an evaluation is contingent upon the accuracy and the truthfulness of the student responses. Correctness also depends on the design of the evaluation instrument and the questions asked.

An interview question was asked about the accuracy of student evaluations of faculty teaching. Professor Betty’s response about accuracy was that she believed students favored certain faculty because of personality. She stated “if they really liked the instructor and they feel he is a good guy they’ll grade that person really high.” She believes accuracy is compromised
because teaching ability is not assessed correctly. She considers instructor temperament and disposition as the major factor in student evaluations.

Professor Betty was also the department chair so she had the opportunity to read all of the department faculty evaluations. She gave an example of a student comment where personality was rated and not teaching ability. She said it was common to read a remark like “Oh, he’s so friendly” instead of something of substance about the professor’s teaching ability. She was dismayed that faculty personality seemed to be more important than teaching acumen or student learning.

Professor Betty also said that evaluations were inaccurate because rater bias exists. She said that faculty were often rated on appearance and students would comment “about the teacher’s attire.” She also lamented that “women don’t know anything about computers and other prejudices shape some student evaluations.” Faculty effort in the classroom doesn’t seem to matter but certain professor qualities like age, personality, and gender do. Her frustration with rater bias caused her to discount evaluations. She experienced other problems with evaluation accuracy beside rater bias.

The correctness of student evaluations can be affected in many ways. Professor Betty illustrated this by articulating that students often use the comment section to vent about non-teaching matters. She said that students have “comments about almost everything including tuition, material fees, textbook fees, and other topics not on the evaluation.” She regrets that students miss an opportunity to evaluate faculty teaching by giving opinions on non-teaching matters and problems. Professor Betty was clearly exasperated with students that do not follow
directions. She seemed angry that evaluations are not accurate as a consequence. The end result is that students missed an opportunity to help faculty improve their pedagogy.

Several participants agreed that students rate faculty based on how well they did in the course. Professor Craig was adamant that an evaluation really had nothing to do with faculty and it had everything to do with a student’s expected mark. He suggested students are really evaluating themselves based on the effort they put into a class. He illustrated this by his comments “those who are getting D’s and F’s because they didn’t show up- would they be able to evaluate the teacher? Gee, they weren’t even there, they didn’t even do the class, those are going to be the bad evaluations.” This participant challenged the notion that evaluations are accurate. He believed that negative reviews are the product of a student that is usually absent. He was clearly angry that the truthfulness of the evaluations was compromised by a student’s poor attendance. His attitude towards student evaluations was bitter. His belief has been constant during his thirty five years of teaching. Poor reviews submitted by someone that is often truant should not count. Professor Craig was frustrated that absentees can affect the evaluation results.

Another faculty member agreed that evaluations were dependent on how the student was doing in the class. Professor Janice suggested that a student that gives a professor a good rating is expecting a good mark in the course. Conversely, a student that is doing poorly in the course and expecting a low grade will give the faculty low marks for their classroom performance. She illustrated this by her comments: I think they tend to rate based on how much they liked the course. Definitely, when students do well in a course they like it more than when they don’t do well in a course. Some of the most well-liked faculty members-they don’t think they are necessarily liked because they educate
better but they have something else about them that students like better. Sometimes it is because they give away a lot of A’s. (Professor Janice, personal communication, May 11, 2016)

Professor Janice’s contention that evaluations are tied to marks has support in various studies. She was displeased about the link between grades and faculty assessments. Her comments illustrated her frustration with evaluations and what she perceived as a lack of useful and accurate information.

Summary. Most participants argued that students measured whether or not they liked the instructor and this contributed to an incorrect evaluation. Two respondents believed that evaluations reflected the expected mark in the class. Other interviewees suggested that assessments weren’t accurate because students had rating bias based on factors like gender, age, race, appearance, and grade expectations. Participants reported that evaluations were not accurate or valid because they did not measure teaching effectiveness.

Significance

Faculty views on student evaluations are a main theme of this investigation and evaluations have two primary purposes. Faculty can take the formative results and make teaching changes. Administration can use the summative results and make key human resource decisions like tenure, raises, and professional development opportunities for faculty.

A common subtheme developed from the interviews was about the importance of student evaluations. Seven out of ten participants felt that evaluations were very significant. Two study subjects said they were not important at all and one participant was unsure about their value.

Professor Janice articulated that evaluations are important. She believes that student feedback is necessary because “the student is the target of what you are doing.” She also stated
that any changes to course design should be student driven. Professor Janice emphasized the importance of student evaluations and how she uses student input to modify and improve her teaching by stating “I think it is important to at least try to get student feedback and make student based decisions when you alter your courses.” She had unwavering support for evaluations but she cautioned that assessments had to contain good information before she could use them for formative purposes. Her enthusiasm was effusive. Her perspective was congruent with many of her colleagues. She was dedicated to improving her teaching skills but she was very frustrated by the lack of good information on assessments. She ascribed to the significance of student evaluations.

The following participant had mixed emotions about the worth of evaluations. She expressed her strong views about their positive value when they are honest assessments. She became agitated when she recalled an example of a bad review. It was not accurate by her account and it represented a missed opportunity for constructive change. Professor Vicky commented:

They mark off items like -does not show up to class on time, doesn’t use the textbook and doesn’t allow for questions. Those are things that I know are not true so I have to throw out the evaluation. So currently I don’t think they are really important in my teaching.

The participant was agitated because she knew those student comments were not true and she had to disregard the evaluation results. She believed an evaluation could only be significant and useful if it was accurate and honest.

Professor Vicky was very animated when she said a false review diminishes the importance of all evaluations to her. She couldn’t use any of the information from an errant review. She believed that an incorrect evaluation was useless. This illustrates the frustration level
that some faculty experience about inaccurate evaluations. This professor normally believed that evaluations were significant but she tempered her remarks by expressing dismay at what some students report. She believes that an incorrect or false evaluation taints the results and decreases its value.

An evaluation may have important information that a professor can use for formative change in the classroom. Professor Joe relished the opportunity to improve his teaching. He reported “if someone doesn’t give you that constructive criticism to change then the individual won’t. I think it is important.” He believes that student evaluations are vital and a catalyst for improvement in the classroom. He suggested that it is easy for an instructor to maintain current teaching methods unless someone gives him a push to amend his lessons and teaching pedagogy. Professor Joe argued that evaluations are opportunities for students to have input into their teaching. Professor Joe said “I feel if the student isn’t given that outlet then they will go more towards the online outlets.” He wasn’t a fan of websites like Rate My Professor.Com. He was resolute that every student’s voice counted and each one should have input into their teaching. He didn’t hide the fact that sometimes he needed something to motivate him to change his pedagogy.

Professor Adam asserted that evaluations are important. He illustrated it by stating: “When I was on tenure track they were important for tenure. Now, they have become something that is actually useful for re-evaluating the class and re-evaluating the pedagogy.” This faculty member had good experiences with student evaluations and contends that they are significant. He said the reason they are important has changed over the years. Good evaluations were useful originally for gaining tenure but now he believes they act as a good measure of his current classroom performance. He appreciates the evaluations for the information they provide and their
potential for formative change. His elation about teaching was evident. He believed in student evaluations and he embraced their potential for positive formative change.

**Outlier.** Professor Eliza suggested that evaluations were not important and perhaps they were potentially harmful. She related a story where administration wanted to hire an adjunct from another department because he had good student evaluations. She was one of two faculty members on the hiring committee. She revealed:

I don’t find them at all important. When you start worrying occurs if administration finds them important. I have actually been in a hearing where our discipline was trying to keep someone from another discipline coming over and it was a competency issue. Administration actually said- Well, he has great student evaluations, doesn’t that sway you at all? I said no because if you pass everyone through they will love you and give you good evaluations. That might have been the first time someone told them that face-to-face. (Professor Eliza, personal communication, May 3, 2016)

Professor Eliza did not mix words-she did not like student evaluations. She was animated during her interview and had a strong conviction against using student assessments to grade faculty classroom performance. She did not find them important and her opinions were scathing at times. She was very passionate in her criticism of student evaluations. Her disapproval was obvious and she shared her delight with the idea that her faculty union was planning to fight for the elimination of student evaluations.

**Summary.** The majority of participants agreed that student evaluations are significant. They believe evaluations represented an opportunity for students to be heard. Assessments are
also important to faculty because they represent junctures for formative change. Any improvement to faculty pedagogy advances student learning and subsequently benefits academia.

Role

A subtheme emerged from the major theme of faculty views on student evaluations and it was the role of faculty evaluations. Ten out of ten faculty members suggested that evaluations should not be the major assessment tool of teaching performance. The study subjects supported this notion by suggesting evaluations should be one part of a multi-faceted approach to assess faculty. Evaluations and their value for faculty assessment were best described by Professor Vicky when she said “I believe it should be an equal mixture of evaluations. Student evaluations shouldn’t be as major as they are here. Definitely peer reviews, Dean visits, we review each other in the department at least once every three years.”

Professor Vicky suggests that student evaluations should be one part of an evaluation plan and they should not play a major role. She bases her valuation on her dislike for student evaluations and what she perceives as a lack of ability by students to measure effective teaching. She affirms that faculty evaluations should complement other methods. She seemed annoyed when she was describing student evaluations. It suggests she had some previous negative experiences with student evaluations.

Professor Adam had a similar perspective. He possessed a hint of anger about the process and the instrument. There was even a suggestion for administrators to go into the classrooms. “How faculty should be evaluated - it should be from personal classroom observation. Administrators, Deans, or whoever needs to go into classrooms and sit there and look and see what happens.” The professor shares the same sentiment as the previous interviewee. They
believe that evaluations should not be the only assessment tool for faculty teaching. They consider evaluations to have a place and a worth but they are only one of many options to assess faculty classroom performance.

Another perspective came from a study participant who was pleased to discuss the role of evaluations. Professor Joe had no complaints. His comments suggested that he thinks evaluations are opportunities to have more contact with his Dean. A lack of communication with the administrator may be a problem not related to the topic of student evaluations. He offers that Dean classroom visits may be an opportunity to have more interaction. “It is important for the Dean to come in so that you can talk to that person on occasion and get constructive criticism.” The professor suggests that student assessments have a value for measuring faculty performance but they should be part of a multifaceted assessment approach. The faculty member illustrates that a Dean’s visit is more valuable to him than a student evaluation. Professor Joe believes that a teaching assessment from his Dean is warranted and preferred to a student evaluation.

Professor Janice agrees with her colleagues that student evaluations should not be the only means to measure classroom performance. She opines that “It could play a role. I would just never say a major role because students don’t necessarily know what helps them learn best.” She also added that “a good evaluation should assist faculty to help their students learn their best.” She conveyed a dislike for student evaluations and how they are currently being used. She did not get evaluated during her first years of teaching at another institution and she regretted the lack of feedback. Her perception was she didn’t get any formative information at a critical early point in her teaching career. She called it “odd maybe and unusual.”
**Outlier.** Professor Craig believed that student evaluations should not play a major role in faculty performance. He laughed as he gave his opinion about his perceived role for student evaluations. He said “if the teacher is competent at all-the negative ones are from those who didn’t do their best in class. They reflected that in their evaluation.” Professor Craig believed that evaluations were not the best way to measure faculty performance so he was adamant that they not be used as the only method to evaluate faculty. He went on to describe his positive evaluation results during his teaching career of 35 years so he didn’t disavow them completely. He was amused with evaluations since his high seniority insulated him from any potential but unlikely trouble the evaluation results might cause him.

**Summary.** Most participants believed that student evaluations had a role in faculty assessment but evaluations should not be the major determinant of teaching performance. Professors suggested that student evaluations could be part of an overall assessment plan that included components like Dean visits, peer reviews and self-assessments. Most faculty members found student evaluations to be essential but they did not want them to be the only measure of faculty performance.

**Process**

A common subtheme developed from the interviews was about the process of student evaluations. A number of factors can affect the process and influence the results of an evaluation. A survey can be paper-based in a classroom or computer generated and completed off-campus. It can be administered by a faculty member or a support person. The survey can be conducted at the end of a semester or concurrently. The results are usually summarized by human resources or
an academic department. A good process is critical for the validity and reliability of student evaluations.

Two questions were asked about the student evaluation process. One question asked if the process was adequate. An example of a faculty response was provided by Professor Vicky. She believed that the process was good but there was always room for improvement.

The process is adequate for the most part. There are guidelines for the timing of the questions but it could be better. The students could be better prepared for why they are being asked the questions and honesty should be stressed.

She appreciated that her school provided guidelines for the timing of evaluations. She also wanted the students to be better prepared and have more truthful responses. Professor Vicky seemed concerned about the process. She says that honesty should be emphasized which suggests that she doesn’t believe that all responses are truthful. She answered the second question about process by referring to the timing of evaluations.

Perhaps student evaluations might be better broken up over the course of the semester. For example, if a student says an instructor is bad even before they’ve been given a test that might carry more weight than at the end of the semester when they may be upset about their grades.

She wants evaluations to be concurrent so that student assessments will not be influenced by their expected marks. She proposes that early and ongoing evaluations might improve the process. It is evident that she has thought about the timing of assessments. Her comments reveal that she found evaluations problematic in the past.

This next comment about the process was made by Professor Betty. She expressed her concern by stating that the current evaluation process could be improved.
I am not certain that the current student evaluation process is adequate given today’s population of students. Although the timing may be appropriate when the evaluation is done near the end of the class, this would not help an instructor who needs to improve immediately. For evaluations given at the conclusion of a course, some students may feel that they have the power to fire or get a teacher fired based upon the comments they write.

She maintained that end of the semester evaluations may be too late for faculty that needed to improve a few months earlier. She was disturbed by the fact that some students believed that evaluations gave them the power to get a professor fired. This professor advocates that evaluations held a couple of times during the semester may benefit both student and faculty. The perception is that she is not supportive of students’ written remarks because some students may write hurtful comments. She seemed to be harboring some resentment towards students that are caustic towards faculty.

Professor Diane did not like the current evaluation process. She stated that “we rarely get feedback in a timely basis and students are concerned about being anonymous. The form is adequate.” This professor suggests that completed forms languish in an administrative office somewhere and are not handed back quickly to the faculty. She was angry about the lack of timeliness of information and the disposition of forms upon completion.

The design of questions is important to the appraisal process. Professor Betty suggested that poorly crafted questions will negatively affect the assessment.

A review of the student evaluation responses and comments might also assist in improving the evaluations. When I read the evaluation comments, I sometimes have noticed that the student comments do not even seem to pertain to the questions asked.
Professor Betty was being frank when she complained about puzzling student responses. The professor says student answers may indicate complex or errant questions. She opines that questions should be regularly reviewed and revised in order to improve the information and the process.

**Summary.** Most participants suggested that the current evaluation process can be improved. Question design, the timing, choosing between end of semester or concurrent evaluations, poor administrator feedback, lack of student training, and puzzling student responses are areas of concern and reason for further study. Improvements to the process will increase the accuracy and credibility of student evaluations.

**Student Training**

A common subtheme derived from the interviews and subsequent data analysis was student training. Six out of ten interviewees believed that students required training before evaluating faculty. One participant said that students could not be taught to assess teaching performance. One respondent did not want the students to be trained because he believes their lack of preparation contributes to more candid answers and strengthens their spur-of-the-moment comments. Two participants opined that student responses and assessments did not require any training.

The following study participants offered comments that are representative of the discussions about student training and its place in student evaluations of faculty. Professor Vicky said that she teaches first year science courses and they are quite different from previous courses the students have taken. She says that the courses are challenging and she holds the students to a high standard. She offers that some students may report that “this is the toughest professor
they’ve ever had.” when in fact she may be the easiest in her department. She believed that students must come into evaluations with awareness. She insists the only way to prepare students to evaluate faculty is with some training beforehand. She suggested “I think it would be nice to have some discussion with the students before their evaluations.” This professor believed that student preparation was a good idea and should be available. She was frustrated with the current process because it wasn’t working for her. She was troubled by the students’ inexperience.

Another participant was philosophical about training. Professor Kevin suggested that the need for training depended on the kinds of questions that are being asked on the evaluation. He said:

If you are asking-is this instructor teaching the material well? Then they may not be properly trained and prepared to know what does that even mean? If the question is-is the grading system fair? That is asking for their perspective. You are not so much looking for a professional evaluation as you are a user evaluation.

Professor Kevin went on to discuss the merits of crafting questions for an evaluation and how question design and choice was critical. He believed that some preparation might help the students. The professor said that training was contingent on the question design.

Professor Craig was asked if students should be trained or not. He argued that training was not required because students evaluated faculty based solely on their anticipated mark. Training would have no bearing on the outcome of the evaluation. He said:

That kind of skirts the question because basically I think now students that put effort into the class like the class. If the student showed up and did the work and the faculty is reasonably competent he will get a pretty good evaluation. The only bad evaluation comes from students
that didn’t try in class. They are evaluating the amount of effort that they themselves put into the class when they fill out a teacher evaluation.

He insisted that students are grading their own work. This participant said it was his long held belief that professors are never actually evaluated. The theory of cognitive consistency suggests that he may be dismissing any negative comments about his teaching because they are inconsistent with his own beliefs.

Another male participant Professor Jordan was asked about training for students and he suggested that while it may be a good idea he believed it really doesn’t impact the end results.

I suppose it would be better. I don’t think it matters. There is no incentive for them—even if I beg them for honest feedback—there is no incentive for them to give it. Theoretically— if there were some points for it.

Students are not forced to provide evaluations and he literally begged them for evaluations. Many students did not provide them. He believed the way to get an evaluation was to award bonus marks for it. Professor Jordan illustrated that training would be a waste of time for students. He went on to discuss Rate My Professor.Com and its merits. He had recently returned to grad school and he had an ineffective teacher in one of his courses. He believed the only legitimate way to evaluate the professor was on the Rate My Professor.Com website. He obviously didn’t have faith in the school’s evaluation process or results. He was mimicking what many of his students did and he did it for the same reasons— he was suspect of the evaluation process and its results. He took matters into his own hands and circumvented the school’s evaluation process. He believed training was not required to get an honest evaluation into the public domain.
Outlier. Professor Eliza suggested that students could not be trained to evaluate faculty. She was visibly angry when she stated “I don’t think you could possibly prepare them to do it. They don’t know what they are looking at.” She explained that students have only one goal and that is to get “through” a class. She said their goal was in conflict with the professor’s notion of what they needed to learn from a class. She was outraged because some professors passed students without any real learning taking place and this set them up for future failure. Professor Eliza was adamant about the training potential of students. She was angry about evaluations and she went on to explain some negative results she had received from past evaluations. She was resentful evaluations were being used. She did not have a problem with other evaluation methods like peer review but she clearly had contempt for student evaluations.

Summary. Most participants in the study supported the idea that students needed training before rating teaching performance. Some respondents cited there was a lack of student knowledge about the meaning of effective teaching. Other study participants concluded that training would be futile because students evaluated faculty based solely on their expected mark or professor likeability and evaluations had no connection to teaching performance.

Student Evaluation Subthemes: Summary. The participants shared some positive and negative perspectives about student evaluations and their accuracy, importance, role, process, and student training in faculty assessment. Many participants challenged the accuracy of student evaluations and affirmed their belief that students rated faculty based on personality, likeability, and whether or not the students enjoyed the course. These ratings conflicted with the perceived main purpose of a faculty evaluation which is to measure teaching performance.
All respondents agreed that student evaluations should not play the key role in faculty performance assessment. Participants were unwavering and passionate in their responses. They were unanimous in their belief that student evaluations needed improvement.

**The Faculty Experience**

Discussions with study subjects generated two key themes and ten subthemes. One of the superordinate themes is the faculty experience with student evaluations. Each one is unique and personal. Student assessments can have an emotional and financial impact on faculty. A positive review can lead to some reward for professors while a negative appraisal can result in remedial training or some form of punishment. Each participant in this inquiry related personal experiences with student evaluations. Interviews and observations generated a number of subthemes. These include opinions, emotions, preparation, formative, and administrative actions.

**Opinions**

A common subtheme illuminated by the interviews with participants was that each study subject had strong opinions about student evaluations. These sentiments were generated from their personal experiences with student evaluations. Every respondent had a private opinion to share. The following quotes and observations capture the subtheme of opinions.

Professor Diane stated that she was encouraged as a novice faculty member when she got “extremely positive” reviews. She expressed it this way: “I got extremely positive evaluations and they encouraged me. I never experienced so much positive feedback in probably anything I had done.” This participant felt student evaluations were a good thing. Her student assessments affirmed for her that she was a good teacher. She had left nursing and entered college teaching. She felt unsure about her standing with the students. She subsequently related that the evaluations energized her and confirmed that she had made the right career choice.
“Students say thank you for all the knowledge you’ve given me. Thank you for all your experience, sharing that with me, you are the best instructor. We loved your class, it was very interesting.” She felt that student evaluations were a win-win for her and her students. She could hardly contain herself when she gave some examples of student comments. Students were effusive in their praise and thanked her for her skill and effort. They stated that they treasured her class and that her lessons were interesting. This faculty member needed affirmation at a critical juncture in her first year of teaching. She opined that student evaluations afforded that opportunity and she benefitted. She did not offer any negative sentiments about student evaluations during our interviews.

Not all participants were benevolent when they gave their opinions on student evaluations. Professor Vicky believed that evaluations didn’t serve their original purpose which was to improve classroom teaching. She lamented their use and said it was difficult to get good information from them. She recounted this example “Half of my students say I’m going too slow and the other half are saying I’m going too fast and I really don’t know what to do with that.” This professor was clearly discouraged because of the mixed messages she received from the students. She didn’t know how to react to those kinds of student comments. Her frustration and bewilderment with the results suggest disappointment with the current evaluation process and diminished expectations for future student assessments.

She offered her perspective that evaluations really didn’t help her personally by stating “You know some of them are good and some say the professor is great. It is nice to hear but again it doesn’t help me.” She didn’t get useful comments to make formative changes to her teaching. She conveyed her concerns by stating that some student responses didn’t offer any
opportunities for improvement. Professor Vicky did not find student evaluations constructive. She believed the information was incomplete and had little value.

Professor Eliza did not like student evaluations and she supported her position by saying that some questions were poorly crafted. She offered “my personal favorite of the questions-I don’t know if they are still on there-did the professor cover all the material? Gee, how would the students know that?”

This participant expressed a sentiment that others articulated. Many believed that evaluations did not provide the information that they were designed for. This study subject seemed upset about the evaluation questions. She suggested later in our interview that the entire exercise was ineffective.

Professor Joe had a strong opinion about student evaluations. He illustrated it by saying: I would have one or two students that didn’t like the class. For the most part, the vast majority- I would ace those evaluations. I noticed that over the last few years I don’t do as well but I don’t believe I’m doing anything differently. Either I have slipped into senility or somehow students have changed. I do have some definite opinions about student evaluations.

He could always count on one or two students that didn’t like his class while the others enjoyed it. He said that his evaluation results were slipping as time passes but he maintains that he hasn’t changed his teaching. He believes that students are changing. The faculty member contends that evaluations aren’t accurate and the results are suspect. He offers his evaluation results as proof. He told me during a later conversation assessments didn’t provide any useful information.
**Outlier.** Professor Adam gave his opinion on the timing of student evaluations. He expressed his belief this way:

I also would assert the need to do evaluations in mid-semester because at the end of the semester you have students that have dropped out or have failed the course or anything like that and they are reacting to a grade as opposed to reacting to teaching.

Students have dropped out or failed and they are grading the professor according to their marks. He asserts that evaluations aren’t measuring faculty performance but are opinion surveys from the students that reflect their satisfaction with their grades. His view was supported by Professor Craig.

**Summary.** All ten interviewees had firm opinions about student evaluations. Four believed evaluations were good because they offered opportunities to improve their teaching. Another participant concurred because her evaluations were always positive and they confirmed her career choice. Two did not like assessments because of inaccurate student comments. Another study subject did not support them because they measured likeability. Two participants did not agree with the evaluation results because they provided confusing information. Faculty opinions about student evaluations matter and should be considered by administration.

**Emotions**

Emotions represent a subtheme of the major theme - the faculty experience with student evaluations. Faculty can feel positive or negative emotions as a consequence of student evaluations. A good review can make a professor happy while a negative assessment may cause a professor to become sad. This investigation and the subsequent analysis of data captured a range of faculty feelings.
Professor Betty gave examples of her positive evaluations by stating “Sometimes they say she’s really easy to understand, the class goes really fast and things like that and it gets you really excited. You think – Okay that worked, it’s really good. You are happy about it.”

It had an effect on her. She said the review was uplifting and it energized her. It also confirmed that her lesson planning and teaching worked. This is an example of a faculty member exhilarated from a good review. She was elated and enjoying the knowledge that her teaching pedagogy is working. She was in the moment and feeling fulfilled as a teacher.

Her euphoria was temporary as she gave me an example of her disappointment and anger with another evaluation. She recounted:

I have 30 students in the class, good scores all the way down except for one person and I say what is with this score here? And you wonder-it’s anonymous – you wonder who is that one? They say she’s not prepared for class and the objectives are not clearly defined.

The participant clearly felt bad about this negative evaluation. She said that a critical review can happen unexpectedly. An errant review must have credence. She was not able to discuss this review with the student because evaluations are anonymous. She was very frustrated.

The next participant had an interesting strategy to deal with the negative emotions derived from a bad review. She offered “A lot of my colleagues just discount negative stuff. They’ll say this faculty just reads her notes and this faculty doesn’t test according to her lecture. Who are they to evaluate faculty?” She questioned the ability of some students to evaluate faculty. This faculty member suggests that evaluations do not affect their self-esteem because many professors discount the comments and do not attach any importance to them. This may be
a coping mechanism for some professors if they do not want to accept any criticism and do not want to make any formative changes.

Professor Eliza had another strategy for negative reviews. She said “the reason I don’t look at them is that they make me angry. When you see students complaining about this or that it is because they do not do their own work and it is very frustrating.”

This participant suggested that she did not take the student evaluations seriously and therefore the results would not affect her self-confidence. Professor Eliza clearly did not trust the student comments and she became angry and frustrated. Ignoring the assessment was her coping strategy.

Not all participants reported that their confidence was intact after their student evaluations. The next account suggests that some professors are upset by student comments and it causes negative emotions and depression. Professor Vicky said:

It can be depressing. I get some classes where the students say disagree, disagree, disagree. It is just depressing. I would rather have a steady emotion at work instead of going up or down. I think they have an overall negative impact.

She recounted how some students will give her poor ratings in every category for no apparent reason. She didn’t like it and she believed the emotional impact was damaging. This professor did not appreciate how evaluations can cause a range of feelings. She did not like disparaging comments. Her anger was evident during our discussion. She also suggested during our interview that Deans focus on negative evaluations and ignore positive student assessments.
The following professor was philosophical about student evaluations and how they might affect mood. Professor Janice illustrates her feelings about student assessments by saying results can be mixed:

I guess it can go either way. When faculty get really good evaluations that can sort of give them positive feedback. It can also go in the other direction where students get a chance to vent and that would make you feel more negative. Actually I get a mix of both- I get some that are really good to read and I’m glad that I read it. There are others that aren’t constructive and it was sort of mean. I get a mix and they are neutral overall.

A positive evaluation can generate good feelings. A negative assessment where the student just vents without a reason can cause sadness or anger. She said that she gets a variety-some positive feedback that she enjoys and some disparaging comments that make her feel bad. Professor Janice says that on balance evaluations for her are neutral. The faculty member experienced good and bad emotions caused by evaluations. She seemed to accept the possibility that evaluations were not going to be positive all the time. She was not emotional during this part of our interview because she was student-focused and she ignored the challenges that evaluations presented. She was frustrated at times by student comments she considered critical or negative.

**Outlier.** Professor Adam had an unusual perspective about student evaluations. He expressed it this way:

I don’t think it really affects morale- for me anyway. Like I said, I like the negative comments. I like negativity so I try to figure out who this was or what this is about as opposed to being depressed-oh I’m a bad teacher. Student evaluations should not affect teaching worth. They should make you reflect on that teaching.
Professor Adam wanted negative reviews. It was surprising to hear. Most faculty prefer positive student reviews. He explained that it really didn’t affect his confidence because he saw negative comments as learning opportunities and a chance for formative change. He thought trying to find out the author of a negative comment was an exercise in futility and it was wasted energy. He refused to let bad comments get him down. He then waxed philosophically about student evaluations. He said “student evaluations should not affect teaching worth. They should make you reflect on that teaching.” This professor’s outlook worked for him and contributed to his positive demeanor.

**Summary.** The results from student evaluations affected the study subjects and caused them to have an emotional response. Interviews revealed that participants felt a range of emotions dependent on the nature of the survey responses from students. The challenge for faculty is to examine those results and recognize formative opportunities. Participants strongly disliked any hurtful or nonsensical comments.

**Preparation**

A subtheme of the faculty experience is preparation. Some studies suggest that faculty change their behaviors/pedagogy to get a favorable rating from students. A question was directed to the participants in this study to ascertain if they altered conduct prior to student evaluations.

The response from Professor Diane was unique and unlike her colleagues’ actions. She said:

I like to bake a lot. I used to make a whole bunch of oatmeal or chocolate chip cookies and if they finish the evaluation we would just give them something for doing it because I wanted the feedback. So years ago I would just make cookies or muffins or something and I would say if you finish it and you know they would be hungry.
She thought it was harmless to bake for the students and didn’t represent any attempt to bribe them for positive evaluations. This faculty member altered her behavior before student evaluations and she maintained that it was entirely innocent. Some might consider it a reward for student contribution while others might consider it an incentive to participate. One could suggest that in each interpretation fresh baking could influence students to assign more positive faculty ratings. During our interview the study subject disavowed any attempt to influence the evaluations. She stated that she was unaware that some faculty try to sway the results of a student assessment.

Faculty members can employ a number of methods to influence students before an evaluation. Some may be perceptible while others may not be. Professor Vicky admitted to being a little nicer before an evaluation but she worried that students may perceive this as contrived. She recalled that “I would maybe be a little nicer but now I am like - what if the students say she is only nice today to get a good evaluation?. I stopped doing that since I’ve moved to the online form.”

Professor Vicky believed that she could impact the evaluation results by improving her manner. She recognized that using online electronic evaluations changed those dynamics. Online forms are perceived to be more confidential and they can be completed anywhere within a specific time frame. Online evaluations are less susceptible to influence by faculty.

Professor Adam appeared very sincere when he said “I tell them the worth of it for myself as an educator. I think that is the most important thing. I do take them to heart.” He says being forthright with the students and identifying the value of evaluations is his only preparation. This participant gave his account of how he handles student evaluations. There is no change in
classroom behavior or any overt action on his part to shape or sway the results. His behavior is consistent with the other study participants. It also illustrates that ethics are important to him.

**Summary.** Eight out of ten participants reported that they have not changed behaviors currently or in the past to influence student evaluations. One study subject said that he would consider it if he was an adjunct and the evaluations impacted whether or not he would be hired full time. The research participants rejected any notion that they would try to influence evaluation results. Some revealed they were not aware of colleagues that tried to influence evaluation results.

**Formative Changes**

A subtheme of the faculty experience is formative change. One of the original purposes of student evaluations was to provide feedback to professors to help them advance their teaching. Formative change occurs when a professor takes a student comment from an evaluation and improves their pedagogy. Professor Vicky’s following comments support the subtheme of formative change.

I use the blackboard often and the students actually write their notes down. I don’t just do PowerPoint. Some of the students were commenting that when I was writing on the blackboard and talking they couldn’t necessarily hear me because the rooms are so large. I would repeat the same thing. They didn’t know this was repetition they just thought it was new material. Now I am just a little more conscious about doing the definition first and then writing on the blackboard and saying – again, this is what this is.

The participant made the required changes because of student comments on the evaluation. A simple change yielded good results for the class. Further probing caused her to
give me another example. “I did have a student write once in an evaluation that with homework assignments they couldn’t keep up with them because of the way they were laid out. I did change that.” The faculty member accepted the changes students had suggested and made the necessary changes. This helped the students and improved her pedagogy. This is an illustration of a participant that wasn’t insulted by negative comments on an evaluation and she didn’t let her ego get in the way of finding a way to assist the students.

Professor Joe recounted an example where he acted on multiple complaints on his evaluations:

One common complaint is that I don’t give enough explanation for things. I usually say-here’s your work, go do it. So now, when I have it up on Desire to Learn which is our platform for online learning I will usually go over the assignment in class which covers 70% of the students or I take the time and make sure if a student has a question to email me.

He admitted to being abrupt with the students and he later confessed that he should have done a better job. He now reviews the assignment in class rather than just leaving some short instructions on the online platform. He also solicits student e-mails after class when someone has a problem. This means that Professor Joe has not let his self-esteem interfere with his student focus. He said the students had a legitimate complaint and he solved it. This professor was open to any positive change in the classroom and he says he was very student focused. He strongly believed that students should have input into their instruction. He used student evaluation comments to make constructive changes.

Professor Kevin used comments from student evaluations to effect change in his teaching style. The following is an example:
I have in the past changed assignment components. I have changed grading components and I have even done things like try to utilize more discussion oriented activities because students have expressed pleasure in the ones that I was using and requested more of them. So I said okay, let’s see how I can work more of these in.

He stated that he improved his pedagogy. He did it for one reason – his students requested it through student evaluations. An interpretation of Professor Kevin’s actions is that he finds evaluations useful and he acts on student comments if they have merit. He wants more than one student to suggest a change before he acts. You could also surmise that he is student focused and wants an interactive learning environment.

He also gave an account of student evaluations he received when teaching online. Almost every student said his assignments are too long and challenging. He accepts this but ignores the calls for a reduction in rigor:

When I teach online almost universally what I hear in my student evaluations is that the course is too much work. You are giving us too much work compared to other online courses I have taken. This is too much work. I am aware of that. I have them work harder than most other online instructors. I do that because most other online instructors softball their students. While I hear that regularly from students for me that is more of a vindication than something that needs to change. That is an example of receiving a comment in the student evaluation that I am going to ignore.

He considers most online courses to be easier than day class so he keeps the difficulty in his course. The faculty member was proud that he challenged the students and he wasn’t going to
allow their complaints to change his pedagogy. He seemed content with how he used student evaluations to make them better students.

**Outlier.** Professor Eliza did not support the use of student evaluations and her irritation was apparent during our interview. Her comments captured her dissatisfaction with student assessments. “I’ve really gotten to the point I haven’t used any information off an evaluation to make any changes because I haven’t gotten any good constructive criticism on them. What I see is not consistent with what is going on.” The professor didn’t like student evaluations and she made that clear. She gave some examples to support her argument. She also made it evident in our interviews that challenging the students was congruent with her teaching philosophy. She stood out among the other research subjects. Student evaluations made her angry and irritated. She couldn’t recount any positive results from evaluations and she had a negative perspective about their value for formative change.

**Summary.** A few professors were combative and ignored formative change suggestions when they received an inaccurate student evaluation. Every participant was willing and anxious to make constructive changes to their pedagogy if they agreed with the comments from the student evaluations. Study subjects agreed that receiving evaluation results in a timely fashion was critical.

**Administrative Action**

A subtheme of the faculty experience is administrative action. Chairs and deans need good information to determine if faculty are effective educators. Most colleges and universities use student evaluations of faculty as their chief measure of teaching performance. Administrators
may use the summative results from student evaluations to make decisions on tenure, raises, and promotions for faculty.

A single question about how administrators might use the results of student evaluations for human resource decisions was posed to faculty during this investigation. The responses varied. Professor Joe recounted a time:

I know when I was up for tenure and went through the process. I submitted all of mine for that time period. There were a couple of bad ones and my discipline chair said your tenure is going to be fine. I just wanted you to know there were a couple of students that were really unhappy with you. I said okay - I wanted to know who. They really don’t sign who. He said to make those changes and other than that just be aware they are out there. It was fine and things worked out for me.

The student concerns did not impact the administration’s decision to grant Joe full tenure. This professor said that administrators used student evaluations as part of the hiring process but they were not the only part. His discipline chair wanted him to address a couple of issues brought forward in the evaluations - not an easy thing to do when you don’t know who the complainants are. The professor also supported the notion that evaluations should only be part of the criteria used for judging teaching performance.

Another faculty, Professor Vicky, did not like the way her former Dean handled student evaluations:

I don’t know how this new Dean is handling the student evaluations. In my experience in the past is that it has been focused too much when they are negative. They haven’t been balanced out when they are positive. In the fall I get bad reviews but in the winter I get good reviews. In the
fall it is because I was a bad teacher. In the winter it is because the students were good. I don’t think it has been fair in the past.

Professor Vicky seemed exasperated at her perceived mistreatment. She seems puzzled at how administrators handle reviews and it contributed to some unnecessary anxiety for her.

Professor Adam was convinced that administrators should not use student evaluations for anything of consequence:

Here I don’t think that they, because of our contract, they cannot use them. They can use them for professional development or suggesting things. They cannot use them as a basis for raises or tenure. That would be a contractual violation.

Professor Adam did not like the idea of administrators using student evaluations as a measure of teaching performance. He believed that teaching should not be judged by students once per semester but rather they should be ongoing and based on “what is happening every week or twice a week.” Professor Adam didn’t like evaluations at all. He didn’t believe they were valid for any purpose. He was firm and demonstrative when we discussed this topic. I didn’t get the sense that he was going to change his perspective any time soon. He worked at a college where the faculty union was strong and not supportive of evaluations. His objections to student assessments may be an indication of past problems with them.

Outlier. Professor Eliza gave her view about student evaluations. She suggested that administrators “should not use them at all.” She also aimed some criticism at administrators without teaching experience when she said “I think administrators that haven’t taught perhaps put a little more stock in them than we do.” Her comments directed at administrators were
succinct. Her words reflected her angst with student evaluations and the current process. Her resentment was reinforced by her support of alternate forms of teaching appraisal.

**Summary.** The majority of participants wanted administrators to use peer reviews, Dean visits, and self-assessments to supplement student evaluations. Many suggested that administrators should regularly visit the classroom to evaluate faculty teaching. The majority of study subjects believed that administrators seemed apathetic about positive evaluations and they focused on negative results. Participants also believed that administration seemed disinterested in evaluations.

**Faculty Experience Subthemes: Summary** Participants shared some positive and negative experiences with student evaluations. They discussed and gave their opinions, described their emotional reactions, and detailed their preparation before student evaluations. The participants also discussed formative changes and their perceptions of the administrative response.

Each study participant had strong opinions about student evaluations. Their emotional reactions varied from happiness to depression. Faculty claimed a lack of preparation for student assessments. All professors were in favor of making formative changes but several claimed that they had never gotten any useful comments on their evaluations. The administrative response drew a variety of negative comments from the faculty.

**Conclusion**

The purpose of this study was to determine the effect of student evaluations on faculty classroom pedagogy. Participants shared their personal stories about student evaluations. A detailed analysis of the interview data and field notes generated an understanding of how faculty
regarded evaluations and in what way the experience affected them. The participant interview data had similarities and was grouped into two superordinate themes and ten subthemes.

The interviewees discussed the merits of student evaluations and provided comments on their accuracy, significance, role, process, and student training. A majority of study subjects stated that student evaluations were not accurate because they believed that students rated likeability and not teaching performance. Most study participants found evaluations to be important but they did not want them to be the only measure of faculty teaching performance. The interviewees suggested that the existing process was inadequate and should be improved. Most faculty in the study believed that students needed training before evaluating teaching performance.

The participants described their experiences with student evaluations. They offered their opinions, described their emotions, explained how they prepare for evaluations, talked about formative changes and discussed the administrative response. Every study subject had strong opinions about evaluations. Participants described their emotions as a consequence of student evaluations and they ranged from joy to sadness. Participants did not change their behaviors to try to influence their students and gain more positive evaluation results. Each study participant was agreeable to make formative changes to their pedagogy if they accepted the comments from the student evaluations. Everyone agreed that administrators should not use student evaluations as the only measure of faculty classroom performance.
Chapter 5: Discussion and Implications for Practice

The purpose of this study was to determine the influence of student evaluations on faculty pedagogy by describing the experiences of certain faculty members at a large urban college. The research question that guided and centered this study was “In what ways and to what extent might faculty compromise good pedagogical practices for positive student evaluations of teaching?” The study was guided by the theory of cognitive consistency (Festinger, 1957) which offers a theoretical viewpoint to help elucidate why faculty behave the way they do before and after student evaluations.

Important information about student evaluations and their effect on faculty pedagogy was collected during this study and supplemented by existing literature. The study results are organized according to the two themes that emerged from analysis of the data: 1. Faculty views of student evaluations, including a. accuracy, b. significance, c. role, d. process, e. student training; 2. Faculty experiences with student evaluations, including a. opinions, b. emotions, c. preparation, d. formative, e. administrative action.

This chapter will present and discuss the findings of this study. The chapter shows how the theoretical framework and literature review supports or challenges the study results. The chapter concludes with a discussion about the findings of the study, recommendations for practice, and its meaning for future research.

Faculty Views on Student Evaluations

A theme derived from an analysis of the study data was the faculty view about student evaluations. Each study participant had a unique perspective and a personal outlook about student evaluations and this contributed to a variety of views about student assessments. Their
observations centered on student evaluations and their perceived accuracy, significance, role, process, and student training requirement. The participants in this study had views that were consistent with the extant literature while in some instances their outlook was contrary to current scholarship.

A sub-theme of faculty views was evaluation accuracy. Participants in this study perceived that student evaluations were not accurate reflections of teaching. Evaluations were designed to measure faculty teaching acumen and the study participants viewed the assessments as inaccurate if the students assessed something other than faculty teaching.

Participants described how students focused on faculty likeability or their expected grades when evaluating a professor. They used terms like “untrue” and “error in rating” when describing their views about student assessment results. A majority of participants also believed that students exhibited rating bias.

Faculty likeability is a primary consideration in student evaluations according to recent scholarship. The literature cites that students focus on the congeniality of professors rather than their teaching and classroom performance. Studies (Huemer, 1998; Winship, 2011) reported faculty know how to increase their popularity and the easiest route to gaining favor with the students is to reduce course difficulty. Ryan et al (1980) found that 38% of professors in one survey admitted to reducing course rigor in an attempt to make studies easier and to appear more likeable.

The research findings are consistent with current literature that reports many students rate faculty based on their expected mark in the class. A student that has done well in a class tends to rate a professor higher that a student that has done poorly. (Huemer, 1998) submits this is a sign
of grading leniency rather than good teaching. The study findings ratify what other studies have found – grades matter to students and faculty.

The study results about inaccurate evaluations complement the current literature on student bias (Griffin, 2001; Morris et al., 1996; Germain & Scandura, 2005). The research findings are consistent with existing studies that report students are influenced by a professor’s appearance, age, race, and gender. Male students tend to rate female professors lower than female students do (Kobitz, 1993). Germain & Scandura (2005) found that minority students tend to rate minority faculty higher. Younger students tend to rate younger professors higher while older students give higher assessments to elder faculty (Germain & Scandura, 2005). Huemer (1998) reported in a study that attractive faculty get better student ratings.

Participants in the study also reported that evaluation results were inaccurate because many students used the general comments section to complain about non-teaching issues like tuition, books, fees, and parking. Study participants considered student evaluations as incorrect if the remarks in the general comments section were not applicable to teaching. Participants felt this represented a missed opportunity to assess faculty pedagogy.

The study findings about accuracy conflicted with some studies that suggested student evaluations were valid and correct. Participants believed that evaluations were measures of professor likeability, expected marks and student bias and not accurate assessments of teaching. Universities and colleges need accurate student appraisals because the formative and summative results are important to the institutions. The faculty view on the significance of evaluations is discussed next.
A subtheme of faculty views is the significance of student evaluations. Assessments of faculty became widespread in the 1960s. Their objective was to grade professors on their teaching and to measure the quality of instruction at universities and colleges (Berk, 2005). Professors can use the evaluation results to make formative changes to their classroom performance. Administrators can make important personnel decisions like raises, promotions, and tenure for faculty based on the summative results of teaching evaluations.

Most participants in this study believed that student evaluations were significant. They used the results for making formative changes to their pedagogy. They offered comments like “useful” and “very important” when they described the significance of student evaluations to their classroom performance. A majority of participants in this investigation accepted student assessments as important and they would consider constructive student comments when they revamped their courses or changed their teaching pedagogy. These views conflicted with participant perceptions about the accuracy of student evaluations of teaching.

The research findings about the significance of student evaluations are supported by current literature. Murray (2005) found that student evaluations had a positive effect on the quality of teaching. He also determined from his research that student evaluations are weighted about 20% - 30% in human resource decisions. He also established in his 1997 study that 73.4% of faculty believed that evaluations provided useful feedback. The results from that same investigation found 68.8% of professors reported that student assessments contributed to an improvement in teaching. Professor Murray has support from many colleagues about the validity and reliability of student evaluations (Wachtel, 1998; Abrami, 2001; Theall & Franklin, 2001).
Some study participants did not find student evaluations to be significant at all because of perceived problems. Their perspectives are consistent with some current literature. Beran & Rokosh (2009) and Winship (2011) found that teaching quality was difficult to measure and evaluations have negligible value. Their findings aligned with the study participants that reported a lack of useful information, challenged the accuracy of the assessment, and detected student bias. This negative judgement about evaluation worth is articulated in recent studies (Huemer, 1998; Fink, 2008; Winship, 2011). Some research participants acknowledged the potential importance of student evaluations but they could not assign any significance when they viewed assessments as incorrect. The next finding in the study indicated some unanimity about the participants’ desired role for student evaluations.

The role of student evaluations was a subtheme of faculty views. Student evaluations have a major role at most colleges and universities (Huemer, 1998). Professors can use the formative results to make changes to their teaching. Administrators can use the summative results to make human resource decisions like tenure and promotion. Scriven (1995) reports that student evaluations are widely used for improving faculty teaching and by administration for personnel decisions. Research supports that students can assess many facets of classroom teaching (Cashin, 1989).

Participants in this study felt that student assessments should play a minor or negligible role at colleges and universities. They noted problems like bias or inaccuracies with evaluations. Most respondents said that students rated faculty based on professor likeability or their expected mark in the class. These findings are congruent with recent scholarship. The literature reports that students give an inaccurate, biased assessment of teaching and evaluations are not valid or reliable (Huemer, 1998; Ory & Ryan, 2001; Simpson & Siguaw, 2000). Study participants
emphasized that evaluations should supplement other assessment methods like peer reviews, dean visits, and self-assessment. The study participants suggested a minor role for student input while noting existing problems with student evaluations.

Alternate assessment methods like peer review were preferred by many of the study subjects and these findings align with current studies. Existing research affirms that peers have the expertise (Cashin, 1989; Sullivan, Buckle, Nicky, & Atkinson, 2012) to provide a thorough and bona fide assessment of teaching inside and outside of the classroom. Peer review can facilitate a mentoring relationship which is beneficial for a new professor.

Many of the participants in this study discussed their preference for teaching portfolios in faculty assessment. These study findings are aligned with current literature. Self-assessment or the use of teaching portfolios is gaining popularity among academics (Centra, 1993; Seldin, Miller, & Seldin, 2010). Faculty gather pertinent information about their teaching and it is then reviewed by the dean or another faculty member.

Some participants stated that they welcomed dean classroom visits to assess their teaching performance. There is a paucity of research about the effectiveness of dean visits but it had the support of some of the research subjects. The next section summarizes the participants request for change in the student evaluation process.

A sub-theme of faculty views was the process of evaluating faculty. The participants in this study characterized the student evaluation process as something that could be improved. The process for student evaluations of faculty at the study location was comparable to the typical evaluation process found at most colleges and universities. Paper evaluation forms are distributed by the professor or a designate at the end of the semester and they are collected and
forwarded to their academic department or human resources for analysis (Layne, DeCristoforo & McGinty, 1999). Some study subjects had concerns about the timing of evaluations. Student evaluations at their institution were mandated for the end of the semester and some participants believed that concurrent evaluations might prove to be more accurate and have more formative value. They reported that concurrent evaluations could address problems early in the semester when faculty members had time to make changes to their lesson plans or pedagogy. The study subjects protested that end of the semester evaluations were too late for teaching adjustments in the current semester and they represented a missed opportunity for an effective teaching change.

Many participants reported that their college student evaluations forms were poorly crafted, contained confusing questions, and gathered too little useful information. Study subjects reported that their departments gave them the latitude to individually create or modify evaluation questions and forms. A few participants reported that they had made changes to their assessment instruments. A few departments had transitioned from paper forms to online forms.

The study findings regarding poor form design aligns with existing literature. Ory & Ryan (2001) found that many assessment forms were created from existing forms at other colleges and universities without consideration “for theory or construct domains” (pg. 32). Evaluations are generally considered poorly designed (Gravestock & Greenleaf, 2008). Kember & Leung (2008) studied questionnaire design and found that only a few surveys were valid or measured what they were supposed to measure.

The study results about concurrent evaluations have support in current literature (McNulty et al, 2010; Keutzer, 1993; Lewis, 2001). A concurrent assessment offers advantages to the students and faculty. Professors can take the evaluation outcomes and amend their
pedagogy or curriculum early in the semester. As a consequence students can feel as if they have some responsibility for course design. The following summarizes the views of the participants on student training needs regarding evaluations.

Student training represents a subtheme of faculty views about student evaluations. A majority of participants felt that students required training to accurately evaluate faculty teaching. Some lamented that student assessments were biased and based on faculty likeability while others offered that student ratings were a product of their expected mark in the class. A majority of participants in this study maintained that training could help students rate faculty based solely on their teaching performance and eliminate other factors.

The need for student training has support in current literature. Training is necessary because students often grade faculty for something other than teaching acumen. Delucchi & Pelowski (2000) found that students rewarded faculty on the student evaluation form for their friendliness. Adams (1997) reported that the perceived approachability of the instructor contributed to better student evaluations. Jameson (2009) went even further and concluded that “their ratings represent their general opinion of the instructor’s acceptability and likeability. The students say so.” (pg. 4) Training would improve evaluation relevance and student rating accuracy.

Some participants said that students required training to change the current practice of assessing faculty based on one’s expected marks. A student will rate a professor higher if they are expecting good marks in a class. Benz & Blatt (1996) found that students took the credit if their grades were satisfactory. They assigned blame to the professor if their marks were not as
good as expected and they rated the faculty lower. Training could ameliorate the problem of evaluating faculty based on the student’s expected mark in class.

The theme of faculty views about student evaluations was supported by the literature as were a number of subthemes including accuracy, significance, role, process, and student training. These views on student evaluations were formed as the result of the study participants’ personal experiences with student assessments. These experiences are discussed in the following section.

**Faculty Experience with Student Evaluations**

A second theme derived from the analysis of the study data was the faculty experience with student evaluations. Each study participant had a personal involvement with student evaluations. These faculty experiences had some common elements and they were designated as subthemes. The subthemes included opinions, emotions, preparation, formative, and administrative action. The findings about the faculty experience were consistent with the extant literature while some experiences were contrary to current scholarship.

A common subtheme resulting from the interviews and ensuing data analysis was faculty opinions. The study participants recounted a number of private stories and shared their opinions about student evaluations.

The majority of participants in this investigation had negative opinions about student evaluations. They did not like student assessments for a variety of reasons. The study subjects supported their opinions with a number of examples. Professors said that students did not have enough information to answer all of the questions correctly on an evaluation. Some students gave puzzling or unkind general comments. Other participants expressed their opinion that students were biased.
Negative faculty opinions about evaluations are reinforced in current literature. Yao & Grady (2005) reported that faculty were anxious at assessment times because they do not like negative feedback. Some study participants reported that bias exists in student evaluations and this is supported by current studies (Germain & Scandura, 2005; Shevlin, Banyard, Davies & Griffiths, 2000; Worthington, 2002). Students may favor professors based on age, gender, attractiveness, humor, race, and likeability. Murray (2005) reported that students lack data to properly assess faculty. They can only rate professors based on observable characteristics and items like course design and instructor knowledge should be left off evaluations.

Some participants gave positive accounts about student evaluations. Their opinions focused on two areas – the comment section of the student evaluation and their overall evaluation score. Positive student remarks generated a sense of pride and fulfilment for some of the study participants. Good assessments meant that they were connecting with the students and their pedagogy was working. Several research participants believed they had made the right career choice and their students had confirmed their decision through their assessments. The study subjects believed that student evaluations were beneficial for academia because they helped to improve their teaching and benefit students.

The positive opinions about student evaluations have support in current literature. Many researchers have opined that evaluations are good measurements of college or university teaching. Beran et al (2005) found in his research that a majority of faculty had positive opinions about student evaluations. Other researchers have concluded that student evaluations are good for assessing effective classroom performance (Abrami, 2001; Marsh, 1987; Schmelkin, Spencer, & Gellman, 1997). Having a strong opinion about student evaluations generates emotions. The study participants’ emotional responses are summarized in the following paragraphs.
A subtheme of the faculty experience was emotions. This study found that evaluations caused an emotional response from the participants. Hargreaves (2001) said “all teaching is inextricably emotional – whether by design or default” (pg. 1057). Existing literature has focused on the cognitive aspects of teaching with scant attention paid to the affective or emotional reaction of faculty in the classroom.

A study by Kogan, Schoenfeld-Tacher, and Hellyer (2010) concluded that female faculty received a more negative impact from student evaluations than their male counterparts. That finding was consistent with this study as a majority of female participants had negative emotions about poor evaluations and used terms like “angry”, “anxious”, “feel bad”, and “affects morale.” The male respondents also received their share of negative reviews. The male emotional response to a bad review was largely neutral. Male professors were usually not disheartened by a negative student evaluation. They generally did not like student evaluations.

The findings of this study with regards to faculty emotions are aligned with current scholarship. Arthur (2009) suggested that a positive rating from students “boosts morale” (pg.448) and makes faculty happy. Participants in this study were glad to receive a positive student assessment. They were exuberant and happy when they described their emotional reactions to a good assessment. Arthur also reports that a negative evaluation can contribute to feelings of “distress, disappointment and hurt” (pg. 448). Participants from this study were generally unhappy or neutral about poor student evaluations.

There are gaps in the current literature about the faculty emotional response to student assessments. The paucity of research about the emotional response of faculty to evaluations may be attributed to academia’s focus on subject knowledge, teaching acumen and student learning
rather than faculty feelings in a diverse and challenging classroom. Negative emotions may cause some study participants to change their behaviors. These behaviors are detailed in the following passages.

A subtheme of the faculty experience is faculty preparation regarding student evaluations. Many faculty will alter their behavior in preparation for student evaluations to receive higher ratings and secure the financial rewards that accompany positive reviews. These behaviors are chronicled in many current studies about student evaluations.

Research has determined that professors will use methods like the recency effect to influence students. They will go into a classroom shortly before evaluations and make an announcement that is beneficial to the students. An example may be to throw out the lowest test score or eliminate some future assignment to gain favor and receive positive evaluations.

Current literature suggests that faculty will dumb down courses and reduce the rigor to influence evaluations. This inflates marks and removes the challenge for students. Students, anticipating higher marks, will grade faculty higher. Arum & Roksa (2010) concluded that this action has reduced the actual learning that takes place at colleges and universities. They attribute grade inflation to faculty knowing how to impact the ratings.

Websites exist (Slater, 2015; McGlynn, 2016; Clement, 2012) that advise faculty how to improve student evaluation results. They tell faculty to dress casually to appear friendlier, use a confident delivery, and be non-confrontational. They also suggest a host of other nurturing behaviors.

The participants in this study did not change their behaviors to try to influence the results of student evaluations. One faculty member said he would consider this action if his employment
was conditional upon positive student evaluations. The participants’ lack of preparation and casual approach to assessment may be attributed to how their institution values student evaluations. Pressure to change behaviors and perform well on student evaluations was absent in this study. The study findings conflicted with current literature that determined some faculty change behaviors in preparation for student evaluations. These new behaviors were designed to manipulate the students and improve assessment scores. The following section describes a positive result of student evaluations: formative change.

One of the original purposes of student evaluations of faculty was formative. Evaluations provided student feedback to professors to help them advance their teaching. The study results align with existing studies where faculty make formative changes to their pedagogy as a result of student evaluations (Stapleton & Murkison, 2001; Simpson & Siguaw, 2000). Faculty that have their pedagogy confirmed by student evaluations would maintain their current teaching practices.

A majority of participants in this study acknowledged using the results of student evaluations for formative purposes. A few study subjects maintained that they had not received any worthwhile information to date from student assessments but they would be open to using any constructive suggestions to improve their teaching.

Participants in this study that used the formative results of evaluations regarded students as capable of judging teaching performance and these study results have support in current literature. Researchers like Murray (2005) have concluded that students can measure effective teaching. Other researchers found that students have the ability to judge faculty performance (Ory & Ryan, 2001; Nasser & Fresko, 2002). Scriven (1995) believes students know if they gained knowledge in a particular course and if the professor has stimulated an interest in the
subject. Students are capable of monitoring faculty behaviors like teaching styles and punctuality (Scrivens, 1995).

Research participants from this investigation that used the formative results from a student evaluation accepted the validity and reliability of the rating instrument. Researchers like d’Apollonia & Abrami (1997) found student evaluation forms to be valid. Student assessment forms have also been found to be reliable (Theall & Franklin, 2001).

A few of the participants in this study said that they did not use the results of student evaluations for formative purposes. They did not believe in the accuracy of evaluations, reported that students exhibited bias, and alleged that students did not have the knowledge or requisite skills to rate teaching performance.

Their rejection of student evaluations and its formative component has support in current scholarship. Winship (2011) reported that colleges and universities emphasize student satisfaction rather than teaching performance in their evaluation instruments. Other research has found that students exhibit bias with regard to instructor age, gender, race, appearance, and likeability of the professor (Balam & Shannon, 2010). The study results aligned with other scholarship regarding the notion that students rate faculty based on their expected grade in the course (Huemer, 1998). Study participants did not always agree about using evaluation results for formative change. They did agree about their experience with administrative action and it is illustrated in the following section.

A common subtheme of the faculty experience is administrative action. Administrative action is prompted by the summative results of student evaluations. Administrators require good information about faculty teaching to make important human resource decisions. Administrators
can take the summative outcomes from student assessments and make decisions about tenure, promotions, hiring, raises, and professional development for faculty.

Participants perceived that administrators did little with the results of student evaluations. The research findings regarding a lack of administrative action were not aligned with current literature. An examination indicated that administrators were atypical when compared to managers at other colleges and universities. The current scholarship about the behavior of administrators regarding student evaluations is:

1. Administrators use student evaluations of faculty as the primary method to evaluate teaching performance (Shao, Anderson, & Newsome, 2007).

2. Administrators use the summative results from evaluations to make major personnel decisions (Beran & Violato, 2005).

3. Administrators are concerned about accountability and good public relations (Murray, 2005).

Full time faculty were assessed at the study college by the students for every course at least once per academic year. These appraisals were not the only method used to evaluate faculty. Faculty also had to provide a written self-appraisal and a peer review to the department dean. Failure to meet deadlines could prompt a classroom visit by the academic dean.

Administrators did not use the summative results of student assessments for personnel decisions. Participants reported that their faculty contract prevented this from happening. Student evaluations were not the major resource for human resource actions like promotion or salary increases.
Respondents reported that the formative benefits from student assessments were diminished because of slow administrator response. Many of the study participants reported that student evaluations sometimes languished in department files and professors did not get timely feedback. The study institution seemed progressive in some aspects of faculty evaluation but acted indifferent with the results.

The Consequences of Student Evaluations

![Diagram](image)

Figure 1: The consequences of student evaluations. K. Hartford (2016)
Conclusion

This study was undertaken to answer the research question “In what ways and to what extent might faculty compromise good pedagogical practices for positive student evaluations of teaching?” The study results found that participants did not change their behaviors or pedagogy in order to receive positive student evaluations.

The study participants did not change teaching activities for a number of reasons. Their college uses multiple assessment methods for faculty evaluation so student assessment was not the only method to appraise academic staff. The study results indicated that there were no financial incentives or rewards for faculty to obtain a positive student rating. Their faculty contract did not include salary increases contingent on good student assessments. Some participants had a disdain for student evaluations. It would be reasonable to conclude that the study participants lacked the desire to influence students to improve their evaluation results when they didn’t validate student evaluation outcomes. Another reason may exist to explain why faculty didn’t change their pedagogy - the participants’ interviews revealed their passion for education. The study participants exhibited a genuine concern for student learning and any compromise to their teaching behaviors would be contradictory.

The study findings are not in alignment with current scholarship in all cases. Some present studies conclude that many faculty at colleges and universities change good pedagogical practices to manipulate students and receive positive student evaluations of teaching. Cahn (1987) concluded that grades largely affect student ratings and students will rate faculty higher in classes where their marks are higher. Faculty have recognized this and changed their behaviors. Ryan et al (1980) discovered that 38% of faculty in their study admitted to dumbing down their
courses in order to influence student evaluations. Fox (2008) lamented that many faculty recognize the path to good evaluations is to reduce the work required of students. Crumbly, Flinn, & Reichelt (2010) concluded that faculty try to entertain students rather than teach them and this has contributed to a decline in course rigor.

Grade inflation has been on the rise since the 1960s (Astin, 1998). Faculty believe that giving out higher marks causes students to reciprocate and return higher faculty ratings (Greenwald, 1997). 25-30% of faculty believe that student evaluations of teaching contribute to grade inflation (Murray, 2005). Faculty can use a variety of methods to influence students in order to get them to return higher student ratings. Simpson & Siguaw (2000) discovered 72 specific behaviors that faculty used to sway evaluation results.

**Recommendations for Practice**

This investigation has important consequences for practice locally and nationwide. The study results provided the researcher with a deeper understanding of the research problem and the process of student assessment. This new comprehension enabled the researcher to formulate an action plan to improve faculty evaluations. A collaboration of students, faculty, and administration will be required for any strategy to effect change with student assessment.

Improvements to student assessments and the evaluation process will benefit academia and society. Faculty will receive accurate assessments of their teaching. Administrators will receive correct summative data and this can be a basis for sound human resource decisions at colleges and universities. Students will receive better instruction and an improvement in real learning. Employers will benefit because graduates are better educated. Society benefits because a better educated populace contributes to a nation’s prosperity.
A crucial first step in the action plan is to promote the merits of improving student evaluations with administration. This initial phase is significant because administration governs colleges and universities and their approval is essential. The researcher can meet one-on-one or with small groups of administrators to discuss the advantages of improved student evaluations. Administration could mandate cooperation from the students and faculty but the best plan would be to gain their support by convincing them about the advantages of improving student evaluations. This could be accomplished by hosting seminars with all constituent groups. The researcher could serve as a facilitator for these meetings. This would be followed by forming representative committees whose task would be to improve student evaluations and process.

The study results were grouped into themes and subthemes. A diagnostic approach to improving evaluations is to examine each subtheme. Not all subthemes will require the full attention of each constituent group. The required methodology is to assign the subthemes to the appropriate member group for a solution. The final product- a new, improved student evaluation and process- could be subject to some sort of vote or approval process to ensure support and compliance from all groups.

Faculty opposition to student ratings has not prompted colleges and universities to noticeably improve evaluation forms or change the process since they first became popular in the 1960s. Current research concludes that higher education is faced with grades being inflated, course rigor being reduced, and actual learning diminished. Many studies blame student evaluations for these troubles. Changes to student assessments of faculty are required if colleges and universities want to correctly measure teaching effectiveness and make appropriate human resource decisions.
The researcher would work with administrators to redefine the role of student assessments. Evaluations must no longer be the primary means to measure teaching quality. Make them one part of a multi-faceted approach to faculty evaluation. Other assessment methods like dean visits, portfolios, peer assessment, and self-assessment are preferred by faculty and have their support.

Student assessments are popular with administrators and they are not being eliminated. Some corrective action needs to be taken as a consequence. The assessment of teaching quality will improve if the accuracy of student evaluations is upgraded. Faculty, students, and administrators must work together to improve question design and process. This will decrease or eliminate student bias and inaccuracies. Academia needs to come up with an industry standard form for evaluations based on an accepted universal definition of effective teaching.

The researcher needs to encourage administrators to develop a training program for students. Students need to be trained to eliminate the unfairness and inaccuracies found in current faculty assessments. Students should earn the right to evaluate faculty by completing a training program. Administrators need to accept the primary responsibility for faculty assessment and stop transferring the entire task for teaching evaluation to the students.

It is essential for faculty and administrators to work together and make all evaluations concurrent. Student assessment at the conclusion of a course is too late. A survey at the end of the semester does not afford the professor the necessary time to make formative changes for that class. The current method of end of the semester surveys may also contribute to students rating professors based on their expected final mark.
Cooperation among students, faculty, and administrators is required to end the confidentiality of student evaluations so that faculty can get more meaningful and honest general comments and issues can be addressed directly with the student. Faculty are currently frustrated because they cannot solve a legitimate problem when they don’t know who the complainant is. Eliminating confidentiality encourages more reflection and serious thought by students and eliminates negative, hurtful comments.

The researcher would convince administrators to stop pressuring faculty to get good student reviews. Many institutions use faculty ratings in their marketing and post these results on their website. This pressure to achieve high ratings may force faculty to compromise their principles and reduce course rigor which increases marks and contributes to grade inflation. Grade inflation and a lack of real student learning is a major problem for academia. The marketing of education implies students are customers and this is inconsistent with the primary goal of education which is to educate students. The researcher needs to convince administrators to eliminate any advertising based on the results of student evaluations.

The researcher must persuade administrators to stop using the summative results from student evaluations as the primary basis for human resource decisions like raises, tenure, and promotions. Administrative visits to the classroom, peer reviews, portfolios or self-assessment offer credible evidence of performance and facilitates human resource decisions. Administrators must be convinced that student evaluation results should not be a major factor in personnel decisions.

The researcher must influence administrators to change the use of student evaluations as student satisfaction surveys. Administrators should accept that evaluations are intended as
formative tools. Students that have legitimate complaints about faculty need to visit the dean’s office and not use evaluations to file a complaint. Faculty members deserve due process and a chance to face their accuser. It works in the justice system and it should work in academia.

Students, faculty, and administrators need to join forces for the benefit of academia and society. Colleges and universities need to recognize that student evaluations have remained relatively unchanged since the 1960s and it is time to bring them into the twenty-first century. Assessment results are often erroneous. A collaborative goal of constructive change to student evaluations is required to improve faculty teaching and increase student learning.

**Recommendations for Future Research**

Further research into the topic of student evaluations and their effect on faculty pedagogy is warranted. A topic for future inquiry is the alternate forms of faculty evaluation. Peer review, dean visits, self-assessment, and portfolios are becoming more popular with faculty (Berk, 2005). These other assessment methods need to be studied to determine if they are feasible alternatives to student evaluations.

Current scholarship places the blame for grade inflation and lower academic standards with faculty (Winship, 2011) and their need for positive student ratings. Grade inflation is real (Germain & Scandura, 2005) and a major problem for academia. Students are graduating without the requisite knowledge required by employers. Administrators want good student retention and high graduation rates. Academia needs to discover if administrators share some of the responsibility for grade inflation.

A related topic worthy of investigation is to determine if students are reducing effort in their studies. Students have reduced study time by 50% over the last 40 years (Winship, 2011).
Less student effort has resulted in a reduction in real learning. Research would help to determine the reason for this decrease in study time and decide if student evaluations are a factor.

Faculty evaluation paper forms are expensive and difficult to administer. Online evaluations are growing in popularity. Research may reveal if online assessments offer genuine benefits like improved accuracy, increased participation, and more timely formative and summative information compared to paper assessments.

Results of this study indicate that students are not experts at evaluating faculty. Students need training if they are to continue assessing faculty. Further study about who should provide the training, how they should be trained, and the perceived benefits of training should be explored. Academia has relied on students to perform a daunting task but has not provided the necessary tools to get the job done correctly.
References


Committee, University of Calgary.


  *Journal for Higher Education Management*, 9(2), 131-139.

  *Assessment & Evaluation in Higher Education*, 35(2), 241-262.
  doi:10.1080/02602930902795927.


Cahn, S. (1987). Faculty members should be evaluated by their peers, not by their students. *Chronicle of Higher Education*, (October 14), B2.


Update (The Newsletter of the Okanagan University College Faculty Association).

Douglas College.


American Psychologist, 52, 1182-1186.


Murdock, B.B. (1962). The serial position effect of free recall. Journal of Experimental...
Psychology 64: 482–488.


Reiners G. (2012) Understanding the Differences between Husserl’s (Descriptive) and


Shao, L., Anderson, L., Newsome, M. (2007). Evaluating teaching effectiveness: Where we are...
and where we should be. *Assessment and Evaluation in Higher Education, 32*, pg. 355-371.


Stapleton, R. & Murkison, G. (2001). Optimizing the fairness of student evaluations: A study of correlations between instructor excellence, study production, learning production, and


Greetings Professor,

My name is Ken Hartford and I am a doctoral student at Northeastern University, Boston, MA. I am in the process of conducting research for my dissertation and I am inviting you to participate in this study.

The topic of my dissertation is “The Effect of Student Evaluations on Faculty Performance.” The first known formal student evaluations of teaching took place in the 1920s at the University of Washington. They are now the primary tool for assessing faculty in the classroom.

Many studies exist concerning how faculty feel about student evaluations. Some professors support their use and submit that they benefit academia. Other studies suggest that evaluations are problematic and lead to grade inflation and a lack of course rigor. This study will examine faculty behavior. Do faculty use the formative results to improve their classroom performance? Might professors try to influence the outcomes by making tests easier in order to get positive student ratings?

This study will include data which will be collected primarily through two semi-structured in-depth interviews. A $15 gift card will be given for your participation.

Participation in this study is voluntary. You may refuse to participate or withdraw at any time. However, if you withdraw before the end of the study, you will receive no payment. Non-participation will not affect your rights or status as a faculty member. You may choose not to answer any questions that you prefer not to.

Please contact only Ken Hartford at: hartford.ke@husky.neu.edu if you wish to participate in this study or you have any questions.

Thank you for your consideration.

Ken Hartford

[Signature]

Schedule A
Signed Informed Consent Document

Northeastern University, Department of Education
Name of Investigator(s): Kimberly Nolan, Ed. D. Principal Investigator; Ken Hartford, Student Researcher
Title of Project: "The Effect of Student Evaluations on Faculty Performance"

Informed Consent to Participate in a Research Study
We are inviting you to take part in a research study. This form will tell you about the study but the researcher will explain it to you first. You may ask this person any questions that you have. When you are ready to make a decision, you may tell the researcher if you want to participate or not. You do not have to participate if you do not want to. If you decide to participate, the researcher will ask you to sign this statement and will give you a copy to keep.

Why am I being asked to take part in this research study?
You are being recruited for this study because you are a tenured professor with at least 3 years teaching experience.

Why is this research study being done?
The purpose of this study is to determine the effect of student evaluations on faculty performance.

What will I be asked to do?
If you decide to take part in this study, we will ask you to meet twice for in-depth interviews.

Where will this take place and how much of my time will it take?
You will be interviewed at a location that you choose. You will be interviewed for a maximum of one hour per session and there should be a maximum of two sessions.

Will there be any risk or discomfort to me?
There is no foreseeable risk or discomfort to you.

Will I benefit by being in this research?
There is no direct benefit to you. The information learned from this study may improve student evaluations of faculty.

Who will see the information about me?
Your part in this study will be confidential. Only the researchers on this study will see the information about you. No reports or publications will use information that can identify you in any way or any individual as being of this project.

Each interview subject will be assigned a number and that will be their identification for the study. The name and assigned number will be retained by the student researcher at his residence under lock and key and nothing else that could identify the participants will be recorded. All data entries and coding will be by the student researcher and only he will have access to this data. The data represents a major source of information for the study so it will be kept permanently in the student researcher’s personal computer that is virus and password protected. Any physical records will be kept in a locked cabinet on a permanent basis and only the student researcher will have access.

APPROVED

Schedule B
NOTIFICATION OF IRB ACTION

Date: March 14, 2016
IRB #: CPS16-02-13

Principal Investigator(s): Kimberly Nolan
Kenneth Hartford

Department: Doctor of Education Program
College of Professional Studies

Address: 20 Belvidere
Northeastern University

Title of Project: The Effect of Student Evaluations on Faculty Performance

Participating Sites: permission forthcoming

DHHS Review Category: Expedited #6, #7

Informed Consents: One (1) signed consent form

Monitoring Interval: 12 months

APPROVAL EXPIRATION DATE: MARCH 13, 2017

INVESTIGATOR’S RESPONSIBILITIES:
1. The informed consent form bearing the IRB approval stamp must be used when recruiting participants into the study.
2. The investigator must notify IRB immediately of unexpected adverse reactions, or new information that may alter our perception of the benefit-risk ratio.
3. Study procedures and files are subject to audit at any time.
4. Any modifications of the protocol or the informed consent as the study progresses must be reviewed and approved by this committee prior to being instituted.
5. Continuing Review Approval for the proposal should be requested at least one month prior to the expiration date above.
6. This approval applies to the protection of human subjects only. It does not apply to any other university approvals that may be necessary.

C. Randall Colvin, Ph.D., Chair
Northeastern University Institutional Review Board

Nan C. Regina, Director
Human Subject Research Protection

Northeastern University FWA #4630
Interview Protocol Form

Faculty Interview Protocol

Location _____________________________________________________________

Interviewee _______________________________________________________

Interviewer _______________________________________________________

Date____________________________ Length in minutes__________________

Age_______________

Educational Level ____________________________
**Introductory Protocol**

My name is Ken Hartford and I am a student researcher conducting a study to understand the effect of student evaluations on faculty pedagogy. Kimberly Nolan, Ed.D. is the Principal Investigator. I would like to audio tape our conversation today. Please sign the release form.

Also note that only the researchers for this study will have access to the audio tapes and they will be destroyed at some point after they have been transcribed. Please sign an additional consent form that meets human subject needs. That form states that your participation is voluntary, you may withdraw from the study at any time, and your risk is negligible. We appreciate your decision to participate.

This interview will be approximately one hour. I have a number of questions that we want to cover and I may prompt you to move forward at certain junctures in order to get answers for all questions.

**Interview Questions**

1. Tell me about your experiences with student evaluations.

   - early in career, now

2. What preparation do students need prior to evaluating faculty?

   - should there be any training? Why or why not?

3. In your experience do students evaluate faculty based solely on performance?

   - please tell me more about that. examples?

4. Tell me about the importance of a student evaluation. Significant or not?

5. Tell me how student evaluations affect faculty morale.

   - how have colleagues reacted? You personally?

Schedule D (2)
6. Tell me how you prepare for student evaluations.

   - do you give the students advance notice?  Are you extra nice that week?

7. Tell me how administrators might use the results of a student evaluation for human resource decisions.

   Raises, tenure, professional development opportunities    Fair or not?

8. How might you change your classroom behavior as a consequence of student evaluations?

   - before?  – after?

9. In your experience should student evaluations play a major role in faculty performance assessment?

   - suggest other ways to evaluate?

10. Tell me how you have used the formative results of a student evaluation?

    - did you change teaching methods willingly?

**Ending protocol**

Thank you for completing this interview. I may have some future follow-up questions after I review the transcripts. I will also share the transcript with you so that you confirm your responses and content. My contact information is hartford.ke@husky.neu.edu if you have any questions about today’s interview or seek more information about this study.

Schedule D (3)
Second Interview Questions

1. Is the current student evaluation process adequate? eg. timing, questions, rating scale, environment

2. Why does your college use student evaluations?

3. Do student evaluations of faculty improve student learning?

4. Can student evaluations be improved?