DISSERTATION

DEFINING AND ASSESSING TEACHING EFFECTIVENESS IN HIGHER EDUCATION

Submitted by

Michele S. Marquitz

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2019

Doctoral Committee:

    Advisor: Carole Makela

    Barbara Wallner
    Paul Shelton
    Jerry Vaske

ABSTRACT

DEFINING AND ASSESSING TEACHING EXCELLENCE IN HIGHER EDUCATION

Teaching effectiveness in higher education is challenging. Given the number of stakeholders and the reasons for assessing teaching effectiveness creates additional challenges. Yet when tying teaching effectiveness to successful student learning outcomes and combining those interests to a case study project, the views of faculty, administrators, and students provided insights and contributed to the body of knowledge of faculty members' performance. Through three manuscripts, we explore defining and assessing a teaching effectiveness process in a case study, using Student Evaluations of Teaching instruments to provide feedback on teaching effectiveness, and the role students' written comments may play in course and instructor feedback.

From analyzing student course surveys to creating qualitative and quantitative instruments with the input of faculty members, teaching effectiveness must ensure successful student learning outcomes. The journey to define and assess teaching effectiveness in higher education was an arduous one presented through three manuscripts. Each manuscript provides insights for new and established faculty members.

The first manuscript presents a case study at a Research I: Doctoral University. Through a research assistantship and partnering with a department challenged to define and assess teaching effectiveness for higher load faculty members, three instruments were developed to determine best practices of effective teachers. The second manuscript used quantitative methods and research to assess students' feedback on faculty members' teaching. And the third

manuscript used qualitative methods to assess themes in written comments from students'

evaluation of teaching surveys.

ACKNOWLEDGEMENTS

It takes a village to accomplish any feat worth undertaking – like a Ph.D. for example. When I began my Ph.D. program, I knew it would be the biggest adventure and challenge I had attempted in my life. While working full-time, my belief in lifelong learning allowed lessons from each class to apply to various aspects of my life. I enjoyed sharing all the theory, insights, and experiences of my classmates with family and friends. How fun that they too could learn about Organizational Learning, Performance, and Change as I was learning. So many family and friends helped me through the good, bad, and crazy times of the doctoral program so I was up for the challenge. I am so grateful for the love and support of my family and friends during this lengthy (and often times, lonely) journey!

Thanks so much to my parents, Daniel and Anita Marquitz, who always instilled in myself, my sister Karen, and my brother Jeff, that we could do anything we set our minds to if we stayed committed. Your love, support, and encouragement throughout the Ph.D. process, though not always expressed by me, meant so much. Thanks also to Karen (and Larry) and Jeff (and Kelly) for your patience and support as well. Those late night, panicked calls when you talked me off the ledge – WOW… Thanks so much! To my nieces and nephew - Ali, Jessie, Livi, Annie, Whitney, and Daniel, I hope you see that lifelong learning is a positive thing. Know that I believe each of you can do anything you put your minds toward as well. You will each make your mark on the world in such wonderful ways in the future.

To Dr. Carole Makela … where do I start with my gratitude for all you have done for me? From the elevation of my writing to pointing me in the right direction for the Research Assistantship, the suggestion to apply to be a consultant at the Writing Center, and so many other

organizations and projects on campus during my doctoral program you shared with me. Thank you so much for your ingenious support groups – from weekly sharing meetings to the writing boot camps to the ability to attend your proposal development class as often as needed for guidance. You helped myself, your advisees, and other students immensely. You have been my greatest advocate and supporter and I am grateful you showed me the way! I am confident I would not be the writer, person, and graduate I am today without your support.

To the rest of my committee, Dr. Barbara Wallner and Dr. Paul Shelton – thank you so much for the support and feedback on my work throughout the process. Dr. Jerry Vaske – a big shout out and thank you for your support as well as I navigated higher education and the dissertation process. Thank you for your willingness to meet and work through some challenging questions and processes. Thanks also for continuing to maintain your passion for your career and students.

It would be silly of me not to thank so many of my fellow Ph.D. students for their constructive feedback, prodding, and constant support. First and foremost, thank you to Dr. Sarah Badding, Dr. Sara Colorosa, and Jerry Zellars (I know you will be Dr. so soon). When we started meeting, through our little group meetings on our own time, who would have thought that we would raise each other up as much as we did and make each of us such better academics? Thank you for making me a better person and student. Our self-created cohort and the work we accomplished together is a gift I will be forever thankful for! To all my other doctoral peers (and there are many of you) who helped provide feedback for my work through Carole's weekly meetings; which made the output for each step of the process better, thanks so much!

My final thank you goes to the amazing tribe of friends I have who have supported me throughout my Ph.D. journey as well as through my life transitions. These friends too have

talked me through the tough times, been patient when I have disappeared for periods of time (working on school work) and have been my biggest cheerleaders. While the list is long of these special friends, thank you so much to Jen and Tim, Jan, Melissa and Steve, Carol and Tim, Joanna and Luis, Barb and Al, Ginny, Ann, Deanna, Amy, Barb, Michelle and Dan, Marge, Donna, Amanda, Peggy and Dana, Jeanette, Meredith and Jack, Stanya and Dave, and so many others! It is difficult to list names of people as there were so many more along the way that supported me, more than they will know; so, a special thank you to anyone I left off the list above!

I am excited to see where my desire for lifelong learning will take me next in this world. With such a strong support system through my family and friends, I am confident I can take on whatever life presents me. I am hanging on and ready for the next adventure!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF KEYWORDS

Teaching Effectiveness

Excellence

Learning

Student Success

Student Evaluation of Teaching (SET)

Continuous Improvement

Faculty Feedback

Quantitative

Qualitative

Whole Person

DEFINITION OF TERMS

In this dissertation, the following terms recognize the definitions listed.

Boyer Commission: a group of individuals, working with Ernest Boyer (then President of The Carnegie Foundation for the Advancement of Teaching), who challenged the status quo in higher education and "proposed four general views of scholarship: discovery, integration, application, and teaching" (Boyer, 1990, p. xii)

Carnegie Foundation: established in 1905, the Carnegie Foundation for the Advancement of Teaching is "committed to developing networks of ideas, individuals, and institutions to advance teaching and learning… They work to advance the discipline of improvement science into education with the goal of building the field's capacity to improve." ("define mission", 1998)

Case study: "research when (1) the main research questions are "how" or "why" questions; (2) a researcher has little or no control over behavioral events; and (3) the focus of the study is contemporary (as opposed to entirely historical) phenomenon" (Yin, 2014, p. 2)

Effective Professor: "one who is intrinsically motivated to learn, because it is he or she who will have the best chance to educate others" (Czikszentmihalyi, 1982, p. 16); someone who successfully transfers the course learning outcomes to students

Engagement: student development and success as gained from the perspective of activities and resources delivery by a course, an instructor, or a department (Holman, 2013)

Formative Assessment: "Formative assessments of teaching are used for the improvement of teaching. They are usually confidential between the instructor and the observer(s) and are

often tied to recommendations for future development activities and opportunities for

improvement." (Dezure, 1999, p. 76)

Learning Outcomes: "identified by the instructor (content, knowledge, or attitudes) that should

be achieved by the students upon completion of the course" (Lieberman, 1999, p. 140),

may be defined by the department; "what faculty want students to know or do as a result

of the instructional experience designed" (Williams, 2015, p. 78)

Pedagogy: an instructional strategy to ensure successful learning outcomes (e.g., checking for

understanding, summarizing, setting objectives, providing feedback, etc.) (Marzano,

Pickering, & Pollock, 2001); Boyer stressed the need to plan appropriate techniques

targeted to particular courses (Boyer, 1990)

Reflective statement (for instructors): introspective review of their work of teaching "calling

attention to their performance in courses, the activities in which they've engaged, and

their contributions to their department's teaching mission" (Palmquist, 2011, p. 6)

Scholarship: "engaging in original research… stepping back from one's investigation, looking

for connections, building bridges between theory and practice, a communicating one's

knowledge effectively to students" (Boyer, 1990, p. 16)

Scholarship of Teaching: the need for faculty to stimulate active learning, excitement for a topic,

to use creativity to transfer information, and to be lifelong learners themselves (Boyer,

1990); allowing teaching to provide a research platform – inquiry to how students learn,

then sharing the results among peers, and building upon the critique(s) provided

(Williams, 2015)

Student Evaluation of Teaching (SET): a feedback method where "…results help to improve the

quality of teaching, as they provide instructors with insight into their strengths and

weaknesses of their teaching practice, based on student opinions" (Spooren, Brockx, &

Mortelmans, 2013, p. 599); formative and summative assessment tool

Student Success: learning has occurred and the faculty member varies techniques to improve

pedagogy for each class (Grassian, 2013); learning transfer has occurred as defined as

"applying previously learned knowledge with various degrees of adaptation or

modification of that knowledge in completing a task or solving problems" (Hung, 2013,

p. 27)

Summative Assessment: review "concerned with providing information to serve decisions or

assist in making judgments…" (Fitzpatrick, Sanders, & Worthen, 2011, p. 21); in the case

of teaching assessment – typically a component of tenure, promotion, or retention

Teaching Effectiveness: a level of performance which takes into account "not only what is

learned by students but also, and importantly, the manner in which the course is designed,

content is selected and delivered, students are engaged in learning activities, and

conditions under which the course is taught (i.e., technology, physical setting, and

students who typically enroll in the course)" (Palmquist, 2011, p. 2)

Teaching Excellence: a level of performance which is "evidenced by high quality teaching

practices (including classroom teaching, assessment practices, module development,

range of teaching, supervision of student projects) and activities in professional self-

development" (Pan et al., 2009, p. 79)

Teaching quality: "the effectiveness with which the teacher is producing the desired learning

outcomes for the given student population" (Weiman, 2015, p. 8)

Tenure: "Freedom of teaching and research and of extramural activities, and a sufficient degree

of economic security to make the profession attractive to men and women of ability"

(AAUP, 1970, p. 14) awarded through a process; impact and protection for the institution as well

Term: length of time a class spans; such as a quarter or semester or a fraction thereof

Workload: amount of effort faculty have designated to teaching, research, and/or service; typically assigned as a percentage of an appointment

CHAPTER 1: OVERVIEW OF STUDY

INTRODUCTION

The scholarship of teaching in higher education listed in the Boyer report states that

teaching is an important focus for faculty, along with the scholarship of discovery, integration,

and application (Boyer, 1990). Boyer believed there was a need for a focus on the scholarship of

teaching with "great teachers creating a common ground of intellectual commitment. They

stimulate active, not passive learning and encourage students to be critical, creative thinkers,

with the capacity to go on learning after their college days are over" (Boyer, 1990, p. 24).

Teaching and learning are words often utilized when discussing student and course outcomes.

There are three additional words, which attempt to address the role faculty play in shaping the

lives of students -- quality, effectiveness, and excellence. Each implies students' learning and

outcomes are contingent on the capabilities and practices of teachers to transfer skills and

learning.

Higher education has a history of change. It continues to evolve with shifts in access to

students, online programs, and loans for students, all impacting faculty to student ratios, diversity

in learning of students, and more. Ironically, Boyer acknowledged nearly 30 years ago

"conditions in higher education have changed significantly in recent years" (Boyer, 1990, p. 1).

Since higher education has become more widely available over the years – land grant

universities, nonprofit colleges, for profit colleges, and online universities, the focus continues to

remain in teaching and student learning outcomes. In addition, the costs associated with these

options have changed with public funding continuing to decrease year after year. There is

concern for the preparedness of freshmen students, relative to the need for remediation or review

instead of moving to the goals of higher education (Carnegie Foundation for the Advancement of

Teaching, 1998; Dezure, 1999). This requires teaching responsibilities to continue to evolve. Faculty members' roles as teachers, guides, coaches, and collaborators are at the forefront of discussions. The Boyer Commission in 1990 expressed these same concerns of change and evolution of teaching effectiveness.

Students represent one stakeholder in higher educations' multitude of systems; there are various challenges to determine what to prioritize per stakeholder (Gray, Froh, & Diamond, 1992; Huber, 2002; Layne, 2012) to ensure learning outcomes. Changes in a system with diverse stakeholders creates opportunities to drive significant improvements by understanding expectations from each of the individuals. It is time to address the question: "How do we recognize and evaluate teaching in higher education?" With the changes occurring, if the goals of higher education are student success, faculty expectations for excellence must continue to evolve and be defined and assessed (Boyer, 1990) particularly in teaching.

There are varied viewpoints on the criteria to determine if successful student learning outcomes are achieved during a course (Beleche, Fairris, & Marks, 2012; Braga, Paccagnella, & Pellizzari, 2014; Cashin, 1999; Grassian, 2013; Huber, 2002; Langbein, 2008; Marincovich, 1999; Palermo, 2013; Stark & Freishtat, 2014). For example, Langbein (2008) posits that students value grades, like to enjoy the courses they spend time and money to take, and want to value learning. Marincovich (1999) argues the need to educate students on their role in evaluating faculty members, which could help increase feedback for faculty members on the status of achieving their course objectives, student success, and learning goals. The varied views on teaching effectiveness above focus on topics, such as who should measure performance, how to measure learning outcomes, and what resources measure students' learning outcomes and effective teaching. Frost and Teodorescu (2001) suggested responsibility for teaching excellence

is with faculty and students are responsible for learning. If we want to utilize the practices of effective faculty members to improve student learning outcomes, how can we learn from effective faculty members' performance and share those practices with other faculty members?

Defining Teaching Effectiveness in Higher Education

Before we can discuss the need to assess faculty members, it is important to align on some of the definitions used in teaching. Many of the educational research articles reviewed lacked a definition for the concepts of teaching quality, effectiveness, or excellence. They only described the concepts in the literature. In 2009, the Pan et al. study identified items students used, positive and negative, to describe teaching based on students' written comments on the Student Evaluation of Teaching surveys (SETs). The study separated the grouped respondents into cohorts based on responses given to teachers' effectiveness on SETs. Descriptors such as interesting, approachable, clarity, ability to explain, effective teaching, and knowledgeable showed as positive indicators in the top 20% cohort list. For the lowest 20% cohort the descriptors were the same except patience was included. Recognizing that many of the definitions of teaching quality only acknowledged traits of the teacher, Weiman (2015) proposed a definition of teaching effectiveness as "the teacher is producing the desired learning outcomes for the given student population" (p. 8). Listing characteristics and performance indicators of effective teaching is more common; though its' assessment continues to remain a challenge.

While defining the items for assessment is difficult, considering the roles of other stakeholders based on their ability to influence the outcome is valuable. For example, administrators influence the achievement of learning outcomes when they assign teaching workloads and courses, distribute classroom assignments, and designate the times of classes (Langbein, 2008). Students' stake is through their levels of engagement (Boyer, 1990; Porter, Rumann, & Pontius, 2011; Rice, 2002), effort (Braga et al., 2014; Huber, 2002), and motivation

to apply new knowledge (Mann, 2010). Students' effort, present or lacking, in a course can also play an integral role in the assessment of faculty; yet again, is challenging to define. Faculty members and stakeholders need clearly defined criteria for assessment.

Universities, colleges, and departments utilize 'codes' to define expectations and provide guidelines on performance. A sample of the sections range from research and community service expectations to selection of faculty, to definition and rationale for tenure. Often there is a performance section in the code detailing expectations or guidelines for teaching, yet the level of detail may not be conducive to guide application by faculty to their teaching. Faculty members need communication of clear definitions, examples, and expectations of satisfactory performance to ensure students' successful learning outcomes. With effective teaching linked to student learning and outcomes, there are numerous other factors contributing to the result (Ried, 2011). Due to the less defined measures and qualitative nature of assessment, defining faculty guidelines continues to be a challenge (North, 1999).

While defining how and what to evaluate for teaching effectiveness, consideration should be focused on defining outcomes of higher education. Faculty members work with students throughout the term. Effective teachers strive to ensure successful learning outcomes in each course. Without clearly articulated expectations and frameworks to evaluate teaching performance consistently, gauging faculty members' performance is a challenge. Some faculty support the idea that student success, as defined by the anticipated learning outcomes for a course, should be the overarching outcome and accountability factor for teaching performance. The Lumina Foundation ("learning outcomes", 2016) defines learning outcomes with active verbs, which students can demonstrate and thus their performance assessed**.** For example, learning outcomes exhibited and assessed, progressively allowing the demonstration of mastery,

reflects faculty members' desire for students to execute the materials. Grades, performance on projects, and preparedness for the next course can reflect successful student learning outcomes. Yet, external characteristics associated with teaching a course as mentioned; such as students' prior knowledge and technology considerations, can influence students' perceptions of whether they learned or not. Clearly articulating expectations for teaching, ensuring successful learning outcomes, and fair evaluation of faculty should lead to student success. Though the particulars of the assessment could vary across departments due to differences in fields, course levels, learning objectives, or other factors, faculty should understand the expectations of their department and institution.

Assessing Teaching Effectiveness in Higher Education

Typically, responsibilities of faculty at Research I: Doctoral Universities are teaching, research, and service. There seems to be more clearly defined expectations on research than teaching performance (Anderson et al., 2011; Carnegie Foundation for the Advancement of Teaching, 1998; Cashin, 1999; Rice, 2002). Yet, teaching allows faculty members direct interaction with students. With these interactions, it can be difficult to assess student learning outcomes. There are various complexities for faculty members when teaching a course, which contribute to the challenges of defining and assessing teaching effectiveness. Assessment in teaching allows for faculty members' input as they may work independently to teach, prepare for class, and grade assignments and assessments. The roles of research and service can, and should be, brought into the classroom and thus used for faculty as a comparison to their peers as well. Boyer (1990) advocated undergraduate students' learning and development occurs by working with faculty on research projects. Active learning and applying learned classroom concepts for students reinforces the importance of the classroom lessons. There are other ways to evaluate teaching, such as through peer reviews. However, a common method of assessment of teaching

is the use of Student Evaluation of Teaching (SET) instruments viewed as one input to the assessment of teaching.

Teaching is a highly personal task. With faculty assessment often relying on the students' input of the "teacher's characteristics and teaching" (Pan et al., 2009, p. 74), do excellent faculty members focus their attention on students' interests and the materials, which need to be communicated? Frost and Teodorescu (2001) ascertain "teaching happens in all interactions between faculty and students" (p. 410). Faculty members' personal beliefs for the best way to convey information to achieve learning outcomes (Grassian, 2013; Layne, 2012) determine their techniques and methods used in a classroom. Czikszentmihalyi (1982) recognized "the real task of a professor is to enable the learner to enjoy learning" (p. 18). His work in intrinsic motivation highlights the outcome of education as students wanting to gain more information on the topic. Yet, various factors can influence students' desires to learn more beyond the faculty members' influence. So, what do excellent faculty members do during a course to ensure student learning? How can assessment of those techniques and methods contribute to the students' learning and the desired course outcomes?

Both formative and summative assessment are forms of evaluation used to assess teaching effectiveness, yet they serve different purposes. Formative assessment focuses on improving the performance of faculty members. While summative assessment represents input in matters of promotion and tenure (Pan et al., 2009). Students can provide feedback relative to formative assessment. Their interaction with faculty members during a term allows them visibility and input to what activities they believed contributed to their learning. However, the scholarly aspect of summative assessment leaves students' competencies lacking. Administrators should help define the purpose of the SETs prior to any administration of the surveys.

Some propose that measuring performance in research, scholarly activity, and service is more objective than teaching due to the quantitative side of research and service. Many academics propose the assessment of research to require the review of research journal ratings, number of articles accepted in peer reviewed journals, number of committees the faculty member participated on, and other quantitative measurements, which can be produced as evidence. Lacking is the quality of the faculty members' contributions in the published research or on the committees where they served. With the rationale of counting articles to assess research, can we utilize the count of classes in a year to represent teaching effectiveness? No! What demonstrates excellent teaching? Are there particular pedagogical methods used by excellent teachers? Are there best practices, by field, in a course used by excellent teachers? What role do student voices play when assessing faculty? These questions point to the differentiation of definitions used when providing feedback on teaching effectiveness and lack of clarity of what effective teachers get their students to do and think. While the literature contains many forms of evidence of teaching effectiveness, such as peer review, self-reflective statements, students' course grades, and students' evaluation of teaching, consensus on how to define and assess teaching effectiveness is lacking. Consideration to the varied uses of these questions and activities within the university setting provides insight to the role of teaching.

Students' ability to assess faculty performance and students' own intellectual skills and knowledge gained from a course are difficult to measure and yet a voice worth capturing. Valid responses relating items on an assessment of learning are difficult for students to gauge due to the possible need for lag time to experience and understand what they learned (Porter et al., 2011). Assessing preparedness for the next course represented by grades may be an indicator of faculty performance (Beleche et al., 2012; Braga et al., 2014) yet difficult to ascertain. Students

may also lack the necessary self-assessment skills (Mann, 2010; McMillan & Hearn, 2008), or be disconnected from the level of learning expected with their own preparedness (Ried, 2011). Many factors, internal and external to students, influence their assessment of faculty members' performance. Identifying the factors which impact the desired learning outcomes for a course can range from prior knowledge of the topic (internal), required versus elective courses (external), time of the class (external), gender of the instructor (external) (MacNeil, Driscoll, & Hunt, 2014), and others (Ried, 2011).

Faculty members have limited control over these factors and SETs may not capture these factors, though some are measurable. Further there is the challenge to attempt to measure the desired students' outcomes, for example, competency in the course content, grades in subsequent courses, or accomplishment in a career. There is a need to ensure the SET design adequately measures indicators of performance. So, what are the best ways for students to provide input that reflects faculty members' practices and students' learning outcomes in a single course?

CASE STUDY DESIGN SELECTION

The teaching effectiveness project (TEP) was an initiative of one department within one college within one university. While the objectives of the project were articulated, the project was a series of requirements and tasks. With an assigned advisory committee, the TEP focused on defining and assessing teaching effectiveness. While entrenched in the project, the teaching effectiveness project became my dissertation topic.

To report and analyze the project, the design chosen was a case study. While the topic of defining and assessing teaching excellence was a relevant and evolving discussion, the realization to analyze an investigation "in depth and within its real-world context" (Yin, 2014, p. 16) was a strong factor in the choice of a case study. This project fit the additional case study definition by Yin (2014) of having "many more variables of interest than data points, reliance on

multiple sources of evidence, and benefits from the prior development of theoretical propositions" (Yin, 2014, p. 17).

In the project when discussing teaching excellence, there were numerous influencing factors -- grades, students' competency, faculty experience, and feedback mechanisms. There was a need to frame specific aspects of teaching to define and assess excellence. Determining the aspects to address relative to teaching effectiveness weighed heavily on the project. An academic environment provided input and multiple sources of evidence.

The schedule for the TEP was two semesters – an aggressive timeline for such a large project. There was the review of research literature, familiarization with the academic world, and immersion in the world of teaching and learning needed at the start of the project. While reading the educational research and determining the data collection needed, as mentioned, I deemed this opportunity for an educational learning experience could serve as the topic for my dissertation. My interest to pursue a career in higher education, fed my desire to learn about teaching, teaching effectiveness, teaching excellence, and academic systems. These lessons would serve a two-fold purpose, data for the project and input for my dissertation.

TEACHING EFFECTIVENESS PROJECT

In fall 2014, a research project began to establish performance standards for new tenure-track positions within a department at Colorado State University (CSU). The approval of a Graduate Research Assistantship (GRA) position in the department allowed the educational research-based project to occur during the 2014-2015 school year.

To support growth in the Masters' programs in the department, new faculty positions were needed and defined with workloads of 70% teaching, 20% research, and 10% service. As positions filled, there was an immediate need to articulate and establish the parameters expected as performance criteria for progress toward tenure and promotion with a specific focus on

teaching loads exceeding 50%. The department code defined tenure track faculty workloads as 50% teaching, 40% research, and 10% service with specific criteria and suggestions listed for research and service. The criteria listed for teaching and thus teaching performance was going to be more difficult to assess. The TEP was underway with the objectives guiding the project: Evaluate and assess teaching performance for progress toward award of tenure, provide voice to faculty in the department on the criteria used to define and assess teaching excellence, identify best practices from teaching excellence research, and define expectations of teaching for faculty with higher teaching workloads. The project focused on the objectives concurrently. The outcome of the project would be to define expectations of teaching when faculty appointments have higher teaching loads – including the defining and assessing criteria for the award of tenure. This remained the goal while researching best practices in teaching effectiveness and providing voice to the departmental faculty throughout the project.

Through informal observations of faculty in staff meetings and with each other, the department faculty members supported the newly formed positions and the TEP. During the departmental retreat for the 2014-2015 academic year, the faculty members appeared to support the department mission, respect each other, discuss differences of opinion respectfully, cooperate with each other, and maintain effective and professional communication. While not everyone spoke during the first meeting, the department exhibited a healthy climate based on Higgerson's (1999) five characteristics of a healthy climate, which would support the TEP and the TEP objectives.

It would be remiss not to mention the challenges for faculty with changes occurring on campuses related to establishing a career in higher education. The American Association of University Professors (AAUP) (2010) published findings on the continued collapse of the faculty infrastructure with tenure track positions declining. Their report found teaching-intensive positions have "risen sharply" (p. 1) over research-intensive positions with "the overwhelming majority of non-tenure-track appointments being teaching only or teaching intensive" (2010, p. 2). Additional challenges surround students and their families weighing in with their expectations for the education experience and desires for classroom experiences (Carnegie Foundation for the Advancement of Teaching, 1998). Expectations of what is taught, how it is taught, and with what resources it is taught no longer follow the "traditional modes" (Carnegie Foundation for the Advancement of Teaching, 1998, p. 1) of teaching; instead the expectations are created based on students and their families. They are stressing the areas they feel are important for education and students' performance. This is not new (Carnegie Foundation for the Advancement of Teaching, 1998) and one area, which is still receiving attention in the press (Yang & Walker, 2015). Another challenge is the shift from federal and state funding directly to institutions of higher education. Many students take on more debt and find alternate ways (e.g., multiple part time jobs and work-study positions) to pay for their education. In a Source Discovery survey of 1,000 adults with children, Yang and Walker (2015) found more loans and other options are available when students are seeking to pay for their education. With all of these changes clearly articulating the expectations of professors and/or administrators is important to ensure the various stakeholders' perspectives recognize teaching excellence.

The definition of teaching effectiveness has been a theme explored on multiple campuses in multiple departments experiencing similar changes similar to the TEP. The Boyer Commission's recommendations and influence on current practices of academics, with increased focus on teaching, played an important role in defining teaching excellence for the TEP. The 25th anniversary of the Boyer Commission's report *Scholarship Reconsidered* (1990) was in 2015. The report challenged the status quo and proposed significant changes in the professoriate at the time. The purpose to trigger discussions "about what faculty members do as scholars on a broad range of fronts" (Rice, 2002, p. 9) succeeded.

In 1990, the original members of the Boyer Commission attempted to define scholarly work or scholarship as a culmination of four areas: "the scholarship of *discovery*; the scholarship of *integration*; the scholarship of *application*; and the scholarship of *teaching*" (Boyer, 1990, p. 16). Scholarship was defined "in earlier times as a variety of creative work carried on in a variety of places, and its integrity was measured by the ability to think, communicate, and learn" (Boyer, 1990, p. 15). The Boyer Commission further defined scholarship as "engaging in original work" (Boyer, 1990, p. 16).

At the 2014 National Conference of the Reinvention Center themed, "Engaged Learning and the Ethos of Discovery – Achieving the Promise in a Tumultuous Era," a panel of four of the original members of the Boyer Commission shared the radical perceptions of their recommendations when released in 1990. The original members of the Boyer Commission at the conference were Shirley Strum Kenny, Bruce Alberts, Charles E. Glassick, and Robert M. O'Neil. Shirley Strum Kenny shared the report, a call to arms, resulted in angry calls from across the country. Many faculty members felt they were already doing what the Boyer Commission suggested or that it would be too expensive to implement. Bruce Alberts received feedback that

the recommendations insulted research universities. And the conversations spurred by the Boyer Commission continue today.

Change is challenging for organizations and the proposed changes of 1990 of increasing the collaboration between faculty and students, suggested by the Commission, met great resistance, more than they anticipated. One example the panel discussed was when they suggested engaging undergraduate students in performing educational research projects. The justification of that proposal highlighted the lessons learned in research could solidify student learning and further faculty members' research platform. This remains contentious today with some educators embracing student involvement in research and others believing it will take too much time and coaching away from the research.

Higher education continues today to strive to engage in original work. These ideas from the Boyer Commission continue to be relevant today in discussions. During the Reinvention Center Conference (2014), University of Texas-Austin faculty shared information about the Bridging Disciplines Program (BDP), founded in 2002, which strives to "support students in becoming versatile thinkers who are able to bring the perspectives, tools, and skills of multiple disciplines to bear on complex issues and questions" ("mission", 2016). With interdisciplinary efforts of students and faculty, each have the opportunity to facilitate hands-on projects to help both students and faculty members. Students have the chance to be stronger stewards of their field gained from the experiences. Faculty members' learnings, personal and through the feedback of the students, can be incorporated into their teaching based on these experiences.

The Council on Undergraduate Research (CUR) founded in 1978, believed "faculty members enhance their teaching and contribution to society by remaining active in research and by involving undergraduates in research" ("about", 2016). The CUR supporters acknowledge

enhanced teaching through working with students. The challenge exists on how to define the purported workloads and then how to measure the benefits to teaching. Many faculty members ascribe to lifelong learning – through both their teaching and their research. At the time of the Boyer Commission in 1990, not terribly different than today at some universities, the negative feedback received about modifying faculty roles to include research with undergraduates revolved around the fear of slowing faculty members' research platforms. Boyer (1990) professed "Theory surely leads to practice. But practice also leads to theory. And teaching at its best, shapes both research and practice" (p. 16).

Educational research continues to be heavily focused on ensuring student learning outcomes are addressed. Universities continue to focus on ways to help faculty succeed in their teaching role by providing various tools (Dezure, 1999; Palmquist, 2011; Ried, 2011; Weiman, 2015; Zubizarreta, 1999), guidance (Frost & Teodorescu, 2001; Mann, 2010), and research suggestions (Carnegie Foundation for the Advancement of Teaching, 1998). Universities have worked to implement the recommendations of the Boyer Commission. Recognized for their efforts and strides in teaching effectiveness at the Reinvention Center conference were University of Texas – Austin and Carnegie Mellon University – Pennsylvania (Reinvention Center proceedings, 2014). There remains opportunities for a larger footprint in academics to engage students in the learning process for the successful transfer of learning outcomes stressed by member organizations of the Reinvention Center.

With multiple initiatives on the campus at CSU, ongoing attempts to address the question "If effective teaching is recognized and evaluated, how are student outcomes assessed" occurred. In 2013, Colorado State University's Faculty and Administrative Professional Manual, section on Performance Expectations for Tenure, Promotion, Merit Salary Increase, specifically teaching

and advising section ("code", 2013) was updated to reflect the focus on teaching by providing suggestions where teaching effectiveness and teaching excellence could be assessed. The suggestions were contained in two paragraphs beginning "Teaching includes…" and "Excellent teachers are characterized by…" which provided examples of items, which assess effective teaching (CSU College code, updated December 2011). There was an additional paragraph, which began "Evaluation criteria of teaching can include, but are not limited to…" A list of inputs for assessing teaching effectiveness and excellence, detailed in Appendix A, remain ambiguous to what excellent performance is when ensuring the achievement of selected outcomes for a course.

In 2011, a task force led by The Institute of Learning and Teaching (TILT) at Colorado State provided input on the assessment of teaching effectiveness. The first three recommendations suggested assessment methods for effective teachers. The fourth recommendation focused on support for faculty members in achieving teaching effectiveness, and thus successful student learning outcomes:

1. Teaching effectiveness should be assessed in part through the use of teaching portfolios during merit, promotion, tenure, and post tenure reviews. The University should develop a web-based portfolio system that will allow faculty members to provide evidence of teaching effectiveness.
2. Teaching effectiveness should also be assessed through peer-observation of teaching.
3. Assessments of teaching effectiveness should include the faculty member's reflective statements on teaching performance and activities.
4. Existing professional development programs supporting teaching effectiveness – in TILT, in the colleges and departments, and in student affairs – should be continued or enhanced. These groups should collaborate on the development of new professional development programs supporting teaching effectiveness. (TILT, 2011, p.1)

In 2014, a new taskforce formed to continue the initial taskforce work published in 2011 on teaching excellence. A diverse group of faculty (TILT, 2015) from across the campus volunteered to participate on the taskforce. The goal of the 2014-2015 task force was to provide

resources to faculty for defining and assessing teaching excellence. Faculty who place a high importance on their skills as teachers, several recognized via distinguished professor awards, voluntarily joined the task force to help assess the issue of teaching effectiveness and successful learning outcomes. The task force targeted concerns over the importance of teaching students and course outcomes (TILT, 2015). The underlying premise of the task force was that excellent teaching is necessary for students to excel in a course and after a course.

CHALLENGES TO DEFINING AND ASSESSING TEACHING EXCELLENCE

The desire to define and assess teaching excellence for the TEP project was an opportunity to learn more about the role of teaching, teaching effectiveness, and student learning outcomes. Dating to the Northwest Ordinance of 1787, encouragement of education systems to promote student learning outcomes has stressed that faculty members are at the center of the system. Given the history of land grant universities with the Land Grant College Act – the Morrill Act of 1862, access to applied education for students' successful use upon graduation, continues to remain a focus. Faculty members often are the ones who individually ties concepts and learnings together for students.

Providing faculty members with practices to use to assess teaching effectiveness is an important step in the success of students. There were other topics related to teacher effectiveness or teacher quality in the research reviewed. However, few provided suggested practices for faculty to implement. Research referenced the challenges with assessing effectiveness or quality of teaching; such as validity of SET scores, reliability of feedback, and competency of students to accurately assess teaching effectiveness. These all fell short of answering "what" and "who" should evaluate faculty performance while acknowledging the reason for the assessment, formative or summative evaluations. The desire still existed for specific criteria faculty members could use to help guide them toward better teaching and expected learning outcomes. The faculty

member and their department establish the course learning objectives, assignments, and assessments, directly impacting students in a course. Faculty members want to see students succeed and many are open to feedback on how their course is achieving the desired learning outcomes. However, what is the best way to assess how a faculty member is performing? Further, who is in the position to assess performance?

Assessment research is a common theme in literature as one input toward evaluation of teaching. Numerous studies reviewed focused on assessment inputs (e.g., student course evaluations, peer review, and portfolios), which can be used to assess both students and faculty members. These criteria need to consider the benefits and challenges of the faculty members' role in teaching and offer various activities and suggested assessment methods while focused on students' learning outcomes.

Given varied faculty workloads of teaching, research, and service, tracking performance progress is important for accountability, ongoing assessment, and continuous improvement. However, there is difficulty in defining workloads, departmental variances in expectations for teaching, research, and service, and tangible measures available to assess faculty members' performance.

Organizational development learnings for the analysis of performance applied to the assessment of faculty members' teaching performance. For example, from Swanson and Holton (2009) the human resource development definition of "a process of developing and unleashing expertise for the purpose of improving, individual, team, work process, and organizational system performance" (p. 4) seemed to align with the TEP. Yet, the complexity of teaching in higher education with faculty members' varied responsibilities produces different results and challenges.

While the goal was to be able to implement performance-based expectations or processes at the individual faculty member level, it was becoming clearer as we progressed through the project, individual performance plans for each faculty member needed to be developed. Without a clearly articulated and accepted job description, it became challenging to discuss assessment or performance in a particular course or group of courses.

ASSUMPTIONS

Teaching is a complex professional endeavor given the variability of students, faculty members, courses, and department values. Some research articles agree students' learning and potential performance in a course, future courses, and careers can be indicators of faculty members' performance (Palermo, 2013; Rice, 2002). Students' performance in a future course or students securing and succeeding in a career can reflect on faculty members' teaching. The number of factors, which could influence student learning, beyond a faculty member's control, are many: term taught, time class scheduled, prerequisite outcomes, classroom setting, students' interest, other students' behaviors, and student learning styles. For the TEP, focus was on the evaluation of teaching for courses in one department. There was no attempt to identify or minimize any influences affecting the effectiveness of a teacher.

There were numerous guiding principles that helped guide the direction of the project. First, the advisory committee members were subject matter experts. They provided needed insights for questions, which arose throughout the TEP. Second, the university, college, and department codes dictated how each level of the organization defined teaching effectiveness or teaching excellence and the suggested inputs to assess teaching. There was a need to align definitions for excellence and assessment of teaching within the university, college, and department codes. Was the code a guideline? What parts were important to faculty or administrators? While the project goal was to ensure evaluative measures for the new teaching

load positions, ensuring all faculty members in the department were striving for excellent teaching quickly became a secondary goal of the project. Finally, there was a need to determine what role students' voices play in assessing faculty performance. While the educational literature spoke to some of these areas, it was necessary to validate collected information with the faculty recognizing perceptions and actual practices may vary within the department and among individuals.

Peers in the department who had successfully exhibited the ability to transfer knowledge and ensure learning outcomes identified the experienced teachers. These experienced teachers acted as role models of behaviors faculty should exhibit for effective teaching. Being able to identify resources to emulate practices successful faculty members had used or the reasons they were able to accomplish their outcomes was significantly more challenging than expected. Supporting the desired outcomes of teaching excellence furthered the discussion between teaching and learning. Students' own the learning aspect in a course per Frost and Teodorescu (2001) yet they rarely 'teach' in the course they are taking. While faculty may utilize group work or group study to enhance student learning, students may not be aware of the techniques and pedagogy or even the purpose of the techniques. However, if students own learning, do faculty members own 'teaching'? Alternatively, do faculty members own teaching in a way that allows students to learn? Moreover, in what situations would this work most effectively?

There are numerous ways to assess teaching excellence. To minimize some of the challenges outside the control of faculty members, teaching excellence assessment focused on factors such as teacher or student motivation, student prior knowledge, career opportunities, and graduate school acceptance. By focusing on performance with a particular course during a particular term, the goal was to identify success factors for student learning outcomes. Yet, other

factors such as relevancy of the material, quality of the classes chosen for the program, and times offered influence a good curriculum or program.

RESEARCHER'S PERSPECTIVE

In the fall of 2014, while pursuing a doctorate part time, my professional career underwent an enormous change as an outcome of an acquisition of the company where I worked. After experiencing numerous acquisitions, system implementations, and reorganizations throughout my 25-year career, this was unexpected. Supply chain operations, my group and my position, moved from the U.S. and Singapore to Penang, Malaysia to newly hired employees. While unexpected and a significant transition, this change provided an opportunity to focus full-time on completing my Ph.D. program. With the preliminary exam completed, focus shifted to the proposal and dissertation stages of my Ph.D. program.

During my education as a Ph.D. student at CSU, my knowledge of the academic structure and hierarchy was limited to my personal classroom experiences. While my professional experiences had taught me to navigate new situations in life, the Ph.D. process challenged me. During my undergraduate work and master's degree, there were set curriculums to follow. Choose a major or program of study and track the courses completed. The Ph.D. program was more difficult. With a professional career that spanned operations and supply chain in the food and beverage, technology, and semi-conductor industries, little prepared me for the overall rigor and isolation of pursuing a doctoral degree. Over time, leading and managing teams in the United States (U.S.) and Asia forced me to expand my comfort zone, acknowledge my personal biases, and to look at situations critically. Each opportunity allowed lessons about new cultures, new ways of managing people, and best practices to push a team to perform better. These same learnings and life lessons helped as I worked through the processes of the Organizational Learning, Performance, and Change doctoral program.

After successfully competing for the Graduate Research Assistant position, based on my work experiences and eagerness to explore the academic research arena, it was time to start working on the TEP. The complexity and challenges of the TEP were unknown at the beginning of the project. With minimal experience in an academic setting, review of the educational research literature and learnings of the experiences of faculty at CSU not only benefited the project; it helped temper my own personal bias and opinions. While there are similarities between higher education and my business world experiences, in defining and assessing learning outcomes, there were limited opportunities to leverage the synergies. The research project highlighted the need for clearly articulated and department agreed upon criteria for assessment of faculty members' performance.

As my dissertation evolved after the project, I expected much of the work to come from the TEP. However, the TEP served as the tip of the iceberg. Two semesters for such a complicated topic served as an 'introductory course' of sorts. It provided a basis of defining and assessing teaching effectiveness in higher education. Concurrently, I realized there were unclear criteria established to define teaching excellence. Grassian (2013) describes the value of utilizing assessment as a looped process revisiting the feedback from the last assessment for its informative value. Grassian (2013) also stresses the need to understand the areas working well and those not working as well and to incorporate that feedback into the teaching practices of faculty members. However, not all feedback provided to faculty serves a purpose. Some may be or may not be constructive, actionable, or applicable to changing assignments. Often, students provide feedback on faculty members' performance through the SET surveys and like myself prior to the TEP, are not aware of the requirements to prepare for a class and/or lack the context of faculty members' workloads, research, service, or academic pressures.

Throughout the TEP, there were multiple attempts to apply years of corporate assessment and performance review experiences. Gauging performance of individual contributors occurs in academics and corporations but the organizational structures of each are different. The premise of promotion also differed between the organizations. For example, tenured professors reap benefits not seen in organizations outside of education; such as, some security of employment, potential need to further research agendas, assurance of a wage to support a level of living, to name a few. The American Association of University Professors (AAUP, 1970) defends two purposes of tenure: first, academic freedom in both research and teaching and secondly, economic security while working in academics. The primary way to discipline a tenured professor is loss of job based on cause accompanied with the option of a hearing (American Association of University Professors, 2010).

There is no job security for individuals succeeding in their job and attaining progress to or hitting their goals, outside of tenured professors. Many states are "at will", which allows the employee or employer to terminate employment at any point for any reason. This represented another variation between higher education and corporate America. The differences between my corporate and academic experiences provided additional opportunities for me to learn.

There had been other opportunities during my years as a student to recognize effective teaching and even some excellent teaching. My definition of a good faculty member included those who were energetic, exhibited passion about their topic, remained current with issues, were fair, provided timely feedback, articulated clear expectations for assignments, utilized creative assignments/lectures, and did not grade too hard. And my definition of a good course included those that were fun, engaging, lecture only, no group work, topics I chose (and liked), and not too hard. In retrospect, several of the items on both of my lists are difficult to clearly define and

assess. Some of the items do not really reflect teaching quality, effectiveness, or excellence. I had learned, applied, and synthesized new concepts in most of the courses and did not use any of those terms when reflecting on teaching effectiveness.

However, these were my opinions as a student and would be difficult to replicate and report on. For example, who decides if the content was current? What if the materials reviewed for the follow on course were not contingent on current issues, should the faculty member try to bring in the current issues? My perspective was typical, as a student – not knowledgeable of pedagogical techniques, preparation of faculty for course work, unsure if this was the first or twentieth time someone taught a class, nor the work associated with each of these factors. I had minimal expectations and knowledge even of my own definition of teaching effectiveness.

Finally, as a student who had completed numerous SETs, I was naïve to the use of the results once I submitted them. As I learned during the TEP, I was quite concerned my peers and other students were equally naïve to the purpose or use of SETs. If the goal is to make faculty members better teachers through feedback, the relevance of the TEP took on a larger role than previously believed. Students play a significant role in the evaluation process and how we try to ensure the evaluation provides 'actionable' feedback for the faculty member is challenging. What is 'actionable' feedback? In addition, what role does feedback play for instructors or in a course? This enlightening moment sent me on a journey to understand how the SETs are useable as an input of the faculty evaluation process.

DELIMITATIONS

The TEP was a nine-month, academic year project focused on establishing promotion and tenure standards for new 70% teaching load, tenure-track positions within one department. The department code defined the expectations of teaching loads at 50%. However, criteria used to assign activities to reflect workload are complex. When defining loads among faculty members,

no time was spent differentiating or defining faculty members' administrative duties, face-to-face courses, online courses, and/or recruitment responsibilities. Since these loads vary by term (and sometimes within a term), expectations were 70% teaching would be simply teaching more than a 50% teaching load.

There was much to learn and accomplish during the project. The goal was to capture consensus from the department head, advisory committee, and department faculty throughout the project, yet there were times when smaller group meetings were required to expedite progress. As much as possible, individual contact with advisory committee members unable to attend the smaller group meetings attempted to maintain alignment within the TEP.

The research undertaken during the TEP concentrated on the effectiveness of faculty in traditional, face-to-face classes. Focusing on these courses allowed the alignment and incorporation of prior research done to define and assess teaching effectiveness. The advisory committee felt the 70% teaching load faculty would spend a larger percentage of their time in face-to-face courses so the related research would be more relevant and applicable.

DISSERTATION FORMAT OF PUBLISHABLE MANUSCRIPTS

This dissertation presents three manuscripts in fulfillment of requirements for the degree of Doctor of Philosophy (Ph.D.). Chapter 1 provides an introduction and literature review. Chapters 2 – 4 each are articles prepared for scholarly journals, each with its own reference list. This format allows the opportunity to gain experiences and competencies in generating educational research for dissemination in targeted, refereed journals. Chapter 5 is a summary of key findings, implications for future research, and reflections.

The first manuscript: A Case Study on Defining and Assessing Teaching Excellence in Higher Education is a descriptive manuscript, which explains the process undertaken in one department at one university to define and assess teaching excellence. Based on the hiring of new

tenure track professors with higher teaching workloads, criteria for success needed clear guidelines. This manuscript describes a department project in 2014-2015 academic year tasked with defining expectations of teaching, best practices from teaching excellence research, providing voice to the departmental faculty members, and finally evaluating and assessing teaching effectiveness for the award of tenure.

Research questions are contained within the remaining two studies. The second manuscript: Student Evaluations of Teaching and Beyond: One Experience – How Will You Use Them is a quantitative analysis of 15,848 records of SET data collected from spring 2011 to fall 2015. The three research questions focused on the SET questions rating the courses and their instructors. The research questions were: 1) What do the descriptive statistics reveal; 2) Is there a difference between lower level courses versus higher level courses; and 3) Is there a difference between male versus female instructors' ratings on SETs?

The final manuscript: Evaluations of Teaching: What We Learn from Students' Written Comments considered the nature of feedback received from students' written comments. The study searched for information faculty members' may gather from the written comments to understand what is or is not working in their courses. This manuscript contained three research questions: (1) How frequently do students respond to the open-ended questions on SETs and provide comments in face-to-face courses; (2) What are the most common comments and how can they inform faculty to ensure successful learning outcomes; and (3) What is the relationship between the written comments and the Likert-rating to the questions rating the course and instructor?

REFERENCES

American Association of University Professors. (1970). 1940 statement of principles on academic freedom and tenure. Retrieved from http://www.aaup.org/report/1940-statement-principles-academic-freedom-and-tenure.

American Association of University Professors. (2010). Tenure and teaching-intensive appointments. Retrieved from http://www.aaup.org/report/tenure-and-teaching-intensive-appointments.

Anderson, W.A., Banerjee, U., Drennan, C.L., Elgin, S.C.R., Epstein, I.R., Handelsman, J., … Warner, I.M. (2011). Changing the culture of science education at research universities. *Science*, *331,* 152-153.

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review*, *31*, 709-719.

Boyer, E.L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate.* New York, NY: The Carnegie Foundation for the Advancement of Teaching.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review, 41*, 71-88.

Carnegie Foundation for the Advancement of Teaching. (1998). *Reinventing undergraduate education: A blueprint for America's research universities.* Princeton, NJ: Boyer Commission on Educating Undergraduates in the research university.

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

Colorado State University Academic Faculty and Administrative Professional Manual (2013). Retrieved from http://facultycouncil.colostate.edu/faculty-manual-section-e/#E.12.1.

Colorado State University – The Institute for Learning and Teaching (TILT) (2011). Evaluating teaching effectiveness task force recommendations. Retrieved from http://tilt.colostate.edu/sotl/taskForces/assessingTeaching/.

Colorado State University – The Institute for Learning and Teaching (TILT) (2015). Evaluating teaching effectiveness task force. Retrieved from https://tilt.colostate.edu/sotl/taskForces/effectiveness/.

Council on Educational Research (CUR).(n.d.). About the Council on Undergraduate Research. Retrieved from www.cur.org/about_cur/

Czikszentmihalyi, M. (1982). Intrinsic motivation and effective teaching: A flow analysis. *New Directions for Teaching and Learning: Motivating Professors to Teach Effectively, 10*, 15-25.

Dezure, D. (1999). Evaluating teaching through peer classroom observation. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 70-96). Bolton, MA: Anker Publishing Company, Inc.

Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines.* Upper Saddle River, NJ: Pearson Education, Inc.

Frost, S.H. & Teodorescu, D. (2001). Teaching excellence: How faculty guided change at a research university. *The Review of Higher Education, 24*(4), 397-415.

Grassian, D. (2013). Constructive ambiguities: The inspiring, deflating, transformative, and limited possibilities for assessment in higher education. *Journal of Excellence in College Teaching, 24*(2), 155-173.

Gray, P.J., Froh, R.C., & Diamond, R.M. (1992). A national study of research universities on the balance between research and undergraduate teaching. *Syracuse University Center for Instructional Development.* 1-23.

Higgerson, M.L. (1999). Building a climate conducive to effective teaching evaluation. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 194-212). Bolton, MA: Anker Publishing Company, Inc.

Holman, K. (2013). *Evaluation of student engagement assessment in Colorado State University's Warner College of Natural Resources*. Retrieved from ProQuest Digital Dissertations.

Huber, M.T. (2002). Faculty evaluation and the development of academic careers. *New Directions for Institutional Research, 114,* 73-83.

Hung, W. (2013). Problem-based learning: A learning environment for enhancing learning transfer. In L.M.R. Kaiser, K. Kaminski, & J.M. Foley (Eds.) *New Directions for Adult and Continuing Education: Learning Transfer in Adult Education, 137,* (pp. 27-38). San Francisco, CA: Jossey-Bass.

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

Layne, L. (2012). Defining effective teaching. *Journal of Excellence in College Teaching, 23(*1), 43-68.

Lieberman, D.A. (1999). Evaluating teaching through electronic assessment. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 139-152). Bolton, MA: Anker Publishing Company, Inc.

Lumina Foundation – Degree Qualification Profile (2016). Retrieved from https://www.luminafoundation.org/files/resources/dqp.pdf

MacNeil, L., Driscoll, A., & Hunt, A.N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40,* 291-303. Published online: 05 December 2014. DOI: 10.1007/s10755-014-9313-4

Mann, K.V. (2010). Self-assessment: The complex process of determining "How we are doing" – A perspective from medical education. *Academy of Management Learning & Education, 9*(2), 305-313.

Marincovich, M. (1999). Using student feedback to improve teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 45-67). Bolton, MA: Anker Publishing Company, Inc.

Marzano, R.J., Pickering, D.J., & Pollock, J.E. (2001). *A Handbook for Classroom Instruction that Works: Research-Based Strategies for Increasing Student Achievement.* Alexandria, VA: ASCD.

McMillan, J.H. & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons, 87*(1), 40-49.

North, J.D. (1999). Administrative courage to evaluate the complexities of teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 183-193). Bolton, MA: Anker Publishing Company, Inc.

Palermo, J. (2013). Linking student evaluations to institutional goals: A change story. *Assessment & Evaluation in Higher Education, 38*(2), 211-223.

Palmquist, M. (2011, December). *University Distinguished Teaching Scholars – The Institute for Learning and Teaching Assessing Teaching Effectiveness Task Force Recommendations.* Retrieved from http://tilt.colostate.edu/sotl/taskforces/teaching/recommendations.pdf.

Pan, D., Tan, G.S.H., Ragupathi, K., Booluck, K., Roop, R., & Ip, Y.K. (2009). Profiling teaching/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Research in Higher Education, 50,* 73-100.

Porter, S.R., Rumann, C., & Pontius J. (2011). The validity of student engagement survey questions: Can we accurately measure academic challenge? *New Directions for Institutional Research, 150,* 87-98.

Reinvention Center (2014). 2014 Conference Proceedings. Engaged learning and the ethos of discovery – achieving the promise in a tumultuous era. Retrieved from http://reinventioncenter.colostate.edu/2014-national-conference-proceedings/.

Rice, R.E. (2002). Beyond *Scholarship Reconsidered:* Toward an enlarged vision of the scholarly work of faculty members. *New Directions for Teaching and Learning, 90,* 7-17.

Ried, L.D. (2011). A model for curricular quality assessment and improvement. *American Journal of Pharmaceutical Education, 75*(10), 1-9.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83(*4), 598-642.

Stark, P.B. & Freishtat, R. (2014). An evaluation of course evaluations. University of California Berkeley. Retrieved from: http://senate.csusb.edu/Reports/evaluations_paper.pdf. 1-15.

Swanson, R.A. & Holton, E.F. (2009). *Foundations of Human Resource Development* (2nd ed.). San Francisco: CA: Berrett-Koehler.

University of Texas Austin (2018). Bridging Disciplines Programs – TEXAS Undergraduate Studies. Retrieved from https://ugs.utexas.edu/bdp/programs.

Weiman, C. (2015). A better way to evaluate undergraduate teaching. *Change.* January/February, 6-15.

Williams, K.M. (2015). *Doing Research to Improve Teaching and Learning: A Guide for College and University Faculty.* New York, NY: Routledge.

Yang, J. & Walker, Z. (2015, Aug 04). Fewer parents paying for college. *USA Today.*

Yin, R.K. (2014). *Case Study Research: Design and Methods – 5th edition.* Thousand Oaks, CA: Sage Publications, Inc.

Zubizarreta, J. (1999). Evaluating teaching through portfolios. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 162-182). Bolton, MA: Anker Publishing Company, Inc.

DEFINING AND ASSESSING TEACHING EXCELLENCE IN HIGHER EDUCATION: A CASE STUDY

SUMMARY

With a continuing need to evolve roles and responsibilities of faculty members at a Research I: Doctoral Universities, a teaching effectiveness project (TEP) began in the fall 2015 semester. The TEP's purpose was to: develop the definition and assessment methods for teaching performance, provide voice to faculty, identify best practices from teaching excellence research, and define expectations of teaching when higher teaching workloads were assigned. Presented as a case study, the development of three instruments occurred during the academic year. The process of developing the interview instrument to capture the voices of faculty and administrators, identifying best practices for effective teachers to ensure successful student learning outcomes, and delivering an additional instrument to allow student feedback were explored. The project's collaboration with faculty and administrators, coupled with research on teaching effectiveness provided much insight into the role of higher teaching load faculty – defining and assessing it. This manuscript presents a framework for taking on such an important task and recounts the steps of the TEP.

INTRODUCTION

The act of teaching is a complex, progressive task. Individuals' perception of what it takes to 'teach' 10-12 hours of coursework a semester is often underestimated and naïve (Cashin, 1999). When discussing the scholarship of teaching, Boyer (1990) noted the "lack of awareness of the hard work and the serious study that undergirds good teaching" (p. 23). From the development of a new course to the execution of an existing course, there are numerous steps in the process which faculty review to ensure relevance of their courses. It requires time to review and to incorporate current events or examples while concurrently adapting to a new student

group each term. Articulating the activities and pedagogy, which underpin the processes faculty members use to develop courses, is more of an exception than the rule (Huber, 2002; Lattuca & Domagal-Goldman, 2007; Layne, 2012). Often the philosophy of the department or university will influence some of the practices. And then there are factors faculty have less influence over: appearance, enthusiasm, energy level (Czikszentmihalyi, 1982; MacNeil, Driscoll, & Hunt, 2014; O'Neill, 1988) and those influenced by the organization or administrators – time class scheduled, technology resources (and their reliability), and physical environment. While using feedback, some from student course surveys, it is possible to update the course content, delivery, and materials. These efforts may be trial and error and may prove effective or ineffective at times in different courses. Developing the content for a course, prior to the actual delivery plays an additional role in teaching effectiveness and the students' learning outcomes.

In higher education the Boyer report (Boyer, 1990) stressed the importance for faculty and universities to focus on the scholarship of teaching, along with the scholarship of discovery, integration, and application. Boyer posited the need to focus on the scholarship of teaching with "great teachers creating a common ground of intellectual commitment. They stimulate active, not passive learning and encourage students to be critical, creative thinkers, with the capacity to go on learning after their college days are over" (Boyer, 1990, p. 24). Educators today have varied roles in higher education. They are cognizant of the areas Boyer stressed with the scholarship of discovery reflected in research. The scholarship of integration represents the collaborative nature of interpreting original research to add to the body of knowledge of teaching. Finally, the scholarship of application moves toward engagement and using faculty's knowledge to apply or further the research via service. Yet the role teaching plays in higher education remains less tangible and more complex to define and assess than the other scholarship areas Boyer defined.

More questions occurred during the initial research than were answered from the literature and advisory group members. For example, when discussing faculty assessment, teaching and learning were words utilized in the literature to discuss student and course outcomes. Three additional words, quality, effectiveness, and excellence, attempted to address the performance faculty play in shaping the lives of students. Each term implies student learning and outcomes are contingent on the capabilities and performance of the teachers to transfer skills and learning. This was the first indicator of the ambiguity of the project. Ensuring the advisory committee aligned on our definition of teaching seemed like an important task, though complicated.

There are varied viewpoints on criteria to effectively determine if course goals (Langbein, 2008, Palmero, 2013; Palmquist, 2011), student success (Frost & Teodorescu, 2001; Grassian, 2013; Palmquist, 2011), and learning (Beleche, Fairris, & Marks, 2012; Braga, Paccagnella, & Pellizzari, 2014; Grassian, 2013; Langbein, 2008; Palermo, 2013; Stark & Freishtat, 2014) are achieved during a course. For example, Langbein (2008) posits that students value grades, like to enjoy the courses they spend time and money to take, and desire to learn. Marincovich (1999) argues the need to educate students on their role in evaluating faculty members, which would help increase feedback for faculty members on the status of achieving their course goals, student success, and learning goals.

The research for the project included topics, such as who should measure, how to measure, and what resources should be used to measure outcomes. Frost and Teodorescu (2001) suggest responsibility for teaching excellence is with faculty while students are responsible for learning. The TEP had a similar question as to where the responsibilities of providing input for assessing faculty members' performance belonged. Do faculty members, administrators,

students, or all of these own providing input on teaching effectiveness. Before answering who should assess and what input, the project needed to determine what to assess!

Given the numerous decisions faculty members make before, during, and after teaching a course, a targeted goal is to ensure achievement of learning outcomes. Mann (2010) defined two types of learning outcomes, affective and cognitive in his study of self-assessment in the medical field. Affective outcomes include reactions, motivations, and self-efficacy. Assessment of students' satisfaction, application of new knowledge, and perceptions of their abilities occur when looking at affective outcomes. Cognitive outcomes include "understanding of task-relevant verbal information, including both factual and skill-based knowledge" (p. 306). Palmquist (2011) and the TILT taskforce (2011) posited "…what students take away from a course in terms of knowledge, skills, attitudes, and abilities – are not synonymous with teaching effectiveness. Although they are closely linked, it is possible (albeit rare) to teach a course well without necessarily achieving the learning outcomes associated with course goals" (p.1). While these ideals seem to contradict each other, it reinforces we are no closer to understanding the impact of teaching effectiveness on student learning outcomes than Boyer in 1990.

Defining the items for assessment to ensure the achievement of students' learning outcomes requires consideration of stakeholders' roles. The stakeholders' ability to influence the outcome is an important factor. For example, administrators have a stake in assessing the achievement of learning outcomes when they coordinate teaching workloads and courses, distribute classroom assignments, and designate the times of classes (Langbein, 2008). Administrators often utilize student course surveys for formative assessment – promotion and tenure decisions. While administrators may be more knowledgeable of the academic trends in higher education, their day-to-day encounters of teaching in the classroom may be limited.

Students' stake in learning outcomes is through their engagement (Boyer, 1990; Porter, Rumann, & Pontius, 2011; Rice, 2002), effort (Braga et al., 2014; Huber, 2002), and motivation to gain new knowledge (Mann, 2010). Students' effort in a course plays an integral role in their learnings, which can influence their assessment of faculty; yet is challenging to define for the purpose of teaching effectiveness. Students' assessment of faculty via the SETs tends to be summative feedback based on their limited knowledge of pedagogy and the academic world. They are able to respond to how they feel the faculty member taught, responded to their questions, and knowledge of the topic in the course. Yet students' competency in any of these areas is limited to their own experience, albeit limited experience.

Faculty members also play a role in ensuring successful student learning outcomes and their own assessment of teaching effectiveness. Research suggests various ways they can identify what is and what is not working in their classes. Peer reviews are a common form of input to assessing teaching effectiveness. There can be a mutual exchange of skills that work and do not work in a course from both individuals' perspective. Faculty members' peers also represent an additional stakeholder invested in helping faculty members succeed.

Another input to assessment, which faculty members can pursue is the comprehensive review of their course materials. Reviewing course materials to ensure clearly articulated learning objectives and assignment requirements can help faculty members realize successful learning outcomes in a course. Miller and Seldin (2016) posit faculty members' written reflective statements on the work they do for a course are additional inputs for assessment. In the reflective statement, faculty identify different pedagogical techniques used and again what worked and did not work. The faculty members' role in assessing their own performance, coupled with students' and administrators' input can help drive successful student learning outcomes.

Research Project Background

In fall 2014, an academic department at a Research I: Doctoral University initiated a project to define and assess teaching. Faculty and staff attended an annual retreat scheduled by the department head to discuss new initiatives for the year and changes that had occurred since the spring semester. Topics included introducing new faculty and staff members, including those members assigned new higher teaching loads. The teaching effectiveness project (TEP) and the departmental advisory board whom had volunteered to contribute to the project were introduced as well. There were questions posed from the beginning of the project by the university and department administrators and faculty about how to provide clear criteria to define and assess teaching effectiveness. It seemed a straightforward project requiring some research, input of faculty and students, and a revised department code on teaching expectations. The sequential steps pursued for the TEP during the academic year of 2014-2015 appear in Figure 1.

The TEP's purpose focused on defining criteria for assessment of effective teachers achieving student learning outcomes. The project included informing promotion and tenure criteria for higher teaching workloads for tenure track, faculty members. The department code had defined tenure track, faculty workloads as 50% teaching, 40% research, and 10% service with specific criteria and suggestions for success listed for research and service ("expectations", n.d.). The department code contained a list of what would qualify as inputs and evidence of teaching effectiveness, such as teacher portfolios, opinions of graduates, and mid-semester course evaluations. Yet there were no details of what a teaching portfolio contained listed in the code. Further, in the tenure and promotion section of the code, evidence of sustained research productivity listed the quantity and quality of research manuscripts based on teaching load. The TEP strove to establish criteria for defining and assessing teaching excellence for the 70% teaching appointment faculty.
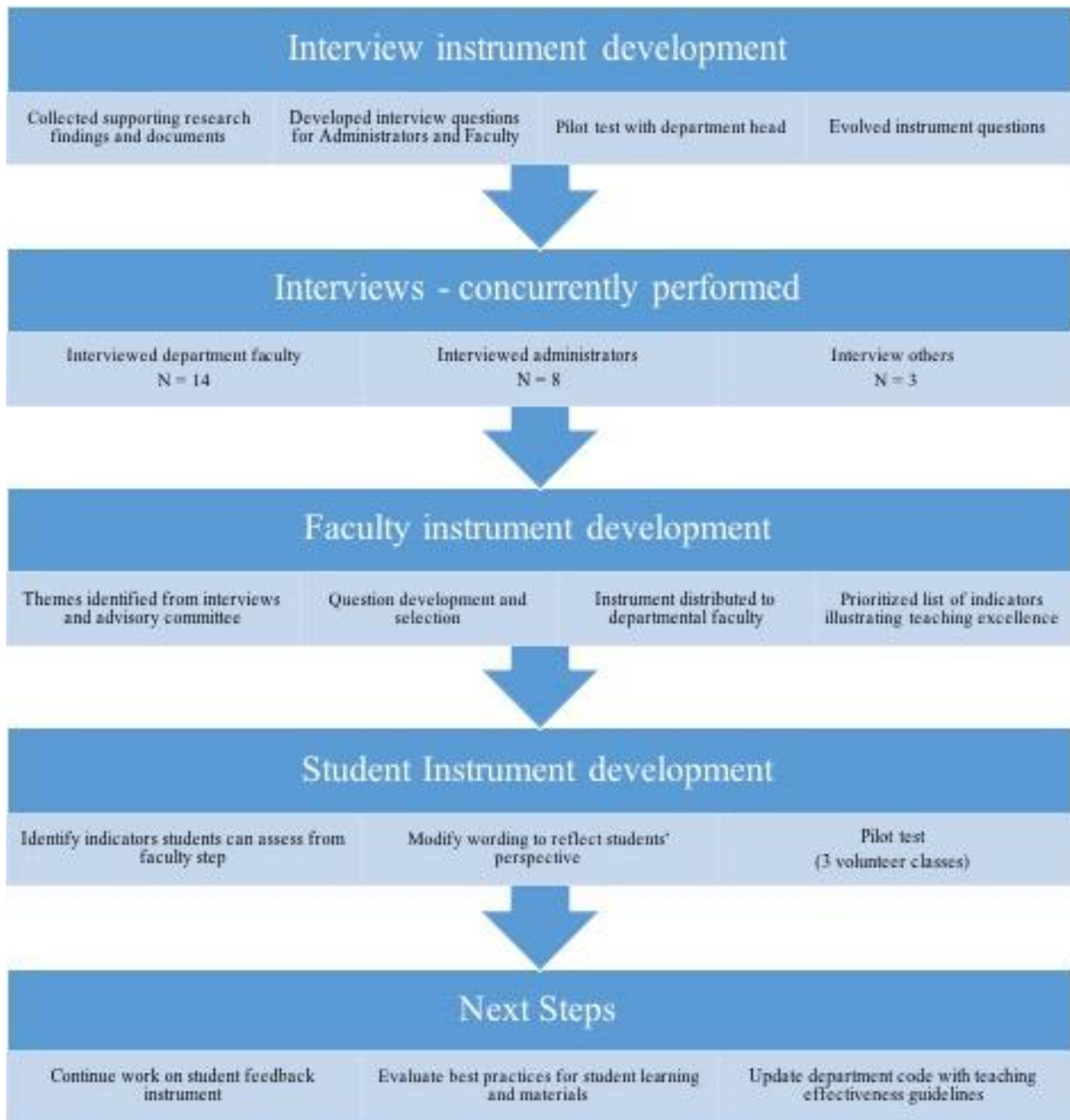
Figure 1. Teaching Effectiveness Project Summary.

The four focus areas listed in the job description for the project defined by the department head are summarized in Figure 2. They were summarized as: (1) to evaluate and assess teaching performance for award of tenure, (2) to provide voice of faculty, (3) to identify best practices from teaching excellence research, and (4) to define expectations of teaching for positions with

higher teaching workloads. The history of teaching in higher education influenced and ensured

past and current research incorporation into the TEP.

- Clarify and improve mechanisms used to evaluate teaching performance in ways that provide constructive feedback to faculty and sustain or enhance excellence in teaching and to develop assessment methods and indicators of teaching success and, scholarship associated with teaching, which will serve as standards for award of tenure.
- Evaluate current situation to identify opportunities for new directions and potential concerns about the initiative among faculty.
- Conduct a broad based assessment of innovative ideas about creating excellence in teaching (e.g., via methods of instruction, ways of organizing, reward systems).
- Delineate in both general and specific the ways to define excellence in teaching performance, teaching scholarship, and scholarship more generally for faculty on higher teaching assignments.

Evaluate and assess teaching performance for award of tenure

Provide voice of faculty

Identify best practices from teaching excellence research

Define expectations of teaching when higher teaching workloads

Figure 2. Teaching Effectiveness Project (TEP) Job Description and Objectives.

LINKING THE PROJECT TO HIGHER EDUCATION

Higher education continues to evolve with shifts in diversity in learning of students,

online formats, changed government funding impacting faculty to student ratios, and access to

distance students' development. Ironically, Boyer acknowledged "Conditions in higher education

have changed significantly in recent years" (Boyer, 1990, p. 1). And the same trend of changes

could be argued as still occurring. Since 1990, higher education access has become more widely

available – land grant universities, nonprofit colleges, for profit colleges, and online universities,

requiring the educational system to adapt to these changes. With these shifts, the role of teaching

and learning in higher education remains a constant focus. Students and parents are taking an

active role in what they expect of the education experience and post-graduation opportunities

(Carnegie Foundation for the Advancement of Teaching, 1998). So many changes are happening.

There are new, evolving opportunities for undergraduate degrees. There is documented concern

on the preparedness of freshmen students, relative to the need for remediation or review. While

faculty want to focus on the goals of higher education (Carnegie Foundation for the

Advancement of Teaching, 1998; Dezure, 1999) often students are not adequately prepared.

Faculty members' roles as teachers, guides, coaches, and collaborators are an important

part of the education process. Faculty members' performance is challenging to assess. There are

responsibilities including but not limited to teaching, research, and service. Preparing for class

and grading assignments and assessments are activities, which provide a unique opportunity to

evaluate students' learning in a course. The impact of teacher and students' desired learning

outcomes while inter-related provide challenges for assessment. Administrators often assign the

courses faculty will teach and where they will teach the course. Yet faculty often have the

flexibility to choose their research topic(s), frequency of research and research format (Langbein,

2008). Faculty members who combine their research with their teaching provide an additional

opportunity for student learning outcomes (Carnegie Foundation for the Advancement of

Teaching, 1998). While these suggestions are helpful to differentiate success in a course, how

can we ensure achievement of teaching and learning outcomes for the students?

Components of Teaching a Course

Teaching requires faculty members to establish and deliver "clear goals, adequate

preparation, appropriate methods, significant results, effective presentation, and reflective

critique" (Huber, 2002, p. 75) in each course. The Colorado State University Administrative and

Professional Manual lists six teaching requirements encompassing effective teaching:

1) Stating clearly the instructional objectives … at the beginning of the term. It is expected faculty will direct their instruction toward the fulfillment of these objectives and that evaluation of student achievement will be consistent with these objectives. Faculty members are responsible for orienting the content of the courses to the published official course descriptions.
2) Informing students of the attendance expectations and consequences, and the methods to be employed in determining the final grade.

3) Assignment of the final grade.
4) Grade examinations, papers, and other sources of evaluation to be available to the student for inspection and discussion. These should be graded promptly to make the results a part of the student's learning experience.
5) Meet their classes regularly and at scheduled times.
6) Make time available for student conferences and advising. Office hours should be convenient to both students and instructor with the opportunity provided for prearranged appointments (CSU university code, 2013, p.16).

While these defined teaching requirements of effective teaching are helpful for both faculty members and students to align on expectations, there is flexibility on the successful completion of each step.

The components of teaching a course as defined by Huber and CSU's University code focused on learning outcomes. Williams (2015) defined learning outcomes as "what faculty want students to know or do as a result of the instructional experience designed" (p. 78). Faculty members determine teaching decisions and desired student learning outcomes. However, once defining the student-centered learning outcomes, faculty need to ensure alignment with the established course learning outcomes. Student-centered learning outcomes are a subset of the students' program of study. Williams (2015) posited faculty members can gauge mastery of the materials taught utilizing students' performance on course work, assignments, and/or exams. Yet Lattuca and Domagal-Goldman (2007) stated any assessment of teaching should include the faculty members' successful ability to achieve the desired student learning outcomes. There are varied opinions among the research as to what encompasses students' learning outcomes, which influences the assessment methods suggested for teaching effectiveness. Interestingly, Grassian (2013) argued "exceptional teaching should not be part of a learning outcome for a course" (p. 167). He went on to highlight that the workplace should have additional input on learning outcomes to ensure the success of graduates.

Evidence of Performance

When it comes to defining or assessing teaching at CSU, the list of expectations listed in the CSU Academic Faculty and Administrative Professional Manual (E.12.1 Teaching and Advising section) (2013) separate teaching and teaching activities. While the list is not all-inclusive (e.g., "… includes, but is not limited to…"), there are references to some of the common activities, such as classroom and/or laboratory instruction and service learning as areas to focus on for teaching effectiveness. The teaching activities listed include traditional responsibilities of teaching such as grading and planning curricula (CSU university code, 2013, p.16). The manual continues to characterize excellent teachers with criteria such as command of subject matter, energy, and enthusiasm. Their guidelines for effective teaching are a start to help faculty members understand the university's expectations. If a faculty member executes the items, listed in the code, is it effective or excellent performance? Some faculty members execute these items effectively based on experience, professional development, or desire to excel in teaching. What differentiates faculty members' performance? With the attempt to define expectations and priorities relevant to faculty members' performance, how do faculty members know they are performing to expectations? A list of factors for the evaluation of teaching (Table 1) are detailed in the university code.

Table 1. Evaluation Criteria of Teaching (CSU university code, p. 46).

Quality of curriculum design
Quality of instructional materials
Achievement of student learning outcomes
Effectiveness at presenting information
Effectiveness at managing class sessions
Encouraging student engagement and critical thinking
Responding to student work

The subjective nature of implementing and evaluating these criteria creates a dilemma for the definition of an effective or excellent teacher. How to evaluate performance remained a question as part of the research project. How should assessment address if faculty members feel they are successfully implementing materials in their course and are able to provide evidence (e.g., course syllabi, letters from students) with no successful learning outcomes achieved? The question continued, who determines if faculty members' are effective teachers.

There are several criteria available to provide evidence of teaching performance. Examples of evidence of excellent teaching are listed in the community code. "Evidence of teaching effectiveness may include…" (CSU university code, n.d., p. 15). These items are "teaching portfolios, professor conducted mid-semester evaluations, evaluations by undergraduate and graduate students of teaching skills, evaluations of other faculty members, letters from students, opinions of graduates, development of new and effective techniques of instruction …" (CSU university code, n.d., p. 15). Similarly, the [departmental] code also highlights similar teaching effectiveness evidence similar to CSU university code as detailed in Table 2.

These criteria do not reflect achievement of the expected learning outcomes in a course; instead, what the faculty member did in the course based on their own assessment or students' feedback. Clearly articulated goals and expectations for teaching performance require as concisely defined criteria as research and/or service. Even with these items available, there are numerous variations of the deliverables provided for them. Often times these are subjective measures of performance for the faculty member, determined by the individual(s) assessing.

Time constraints needed to implement and deliver the evidence plays an additional role in faculty members' ability to produce some of this evidence. Some evidence requires less time to

gather, such as letters from students, versus the expectations of deliverables, such as development of new and effective techniques of instruction. Thus, what influences faculty members' uses of these evidence: ease of implementation, time required to complete, or departmental acceptance of items?

Table 2. Examples of Evidence of Effective Teaching (CSU university code, p. 47).

Faculty Provided	Peer/Student Provided
Course syllabi	Signed peer evaluations
Examples of course improvements made	Letters, electronic mail message, and/or other forms of written comments from current and/or former students
Integration of service learning	Appropriate course surveys of teaching
Professional development related to teaching and learning	
Development of new courses and teaching techniques	
Student learning achievement	
Evidence of the use of active and/or experiential learning	
Assessments from conference/workshop attendees	

The opportunity for improvement occurs in different aspects of life and teaching is no exception. Implementing change or continuing what a faculty member is doing, based on student, peer, and administrator feedback allows faculty members' participation in the process. The performance driving excellent teaching is a continuum with numerous factors influencing it. The university code recognizes some of the factors, which may influence faculty members' performance. The code references the "physical and curricular context in which teaching occurs". It further offers examples, which differentiate teaching techniques; such as level of course and course delivery method, which influence excellent teaching. Few would argue that face-to-face classes utilize different pedagogical techniques or offer different opportunities from online classes to ensure successful learning outcomes.

Ensuring there is a healthy climate in a department for the effective assessment of teaching is critical to drive improvement in courses (Higgerson, 1999). Assessment should not be perceived as a formative activity of teaching for the improvement of an individual's work (Cashin, 1999). Instead, assessment represents constructive feedback of what is going well in a course and opportunities for improvement. Documentation of evidence of performance could include activities such as the review of course materials, peer review of course delivery and materials, student course evaluations, faculty reflective statements, and other identified assessment tools from the literature. How do we know if the achievement of students' learning and the completion and attainment of the learning outcomes occurred?

Quality in Teaching

Much work goes into teaching and keeping in perspective the outcome of the student learning experience. The work faculty members do prior to the start of a course remains an important perspective to capture. To determine faculty members' level of teaching excellence from the perspective of students, there are various activities and pedagogical techniques to incorporate into a course.

There are positive contributors to learning identified in the literature with standard criteria of activities and outcomes. The research details the value of achieving learning outcomes and students' experiences when incorporating active learning, student engagement, critical thinking skills, writing skills, and problem solving skills into courses. The research challenges who owns the activities and which of the pedagogical techniques contribute most to the success of learning outcomes. Frost and Teodorescu (2001) posited the faculty member **and** the students each own the responsibility for ensuring successful learning outcomes. Acknowledging and sharing the learning outcomes in a course addresses ownership when assessing faculty members'

performance. Both faculty members and students play an active role in achieving student learning outcomes.

Most courses require faculty member updates from term to term: syllabus, incorporation of current events (when applicable), updating materials utilized in the last session, and other tasks. Once the term starts, there are additional activities faculty members encounter necessary for the course. Some of these activities are tackling technology requirements, preparing lesson plans, creating rubrics, explaining project details, and grading assignments. Further many faculty members make time to interact and get to know students – face to face, e-mails, scheduled appointments, or with the use of technology. All of these activities become time demands outside of the hours spent in the classroom. Most of these tasks are performed weekly and considered part of teaching a course. As Lattuca and Domagal-Goldman (2007) noted "… the individual instructor is a member of a large program or department with a collective goal of educating students in an academic domain broader than the classroom." (p. 15). While initially requiring effort and time, the result for the students' experiences should result in better practices among faculty over time (Spooren, Brockx, & Mortelmans, 2013).

Student Evaluation of Teaching (SET)

Many teaching surveys attempt to ask students questions about a course and instructor in an attempt to gauge teaching effectiveness. Questions are associated with factors discussed thus far, which are essential to students' learning. Often, there may be a question about student learning. For example, as part of the CSU Student Course Survey there is a question "How effectively did the instructor facilitate student learning?" This question is difficult to assess, as the students may not compare their learning relative to the stated course goals or the instructors' role (Palmquist, 2011). With so many factors influencing the ability of the faculty member to facilitate the desired learning outcomes, this question represents only one aspect of the course.

Assuming SET design is reliable and matched to a particular purpose; students often are not in a position to assess the stated items (Cashin, 1999; Layne, 2012; Porter et al., 2011; Stark & Freishtat, 2014). A recent experiment demonstrates the lack of students' ability to assess faculty members' performance instead providing an opinion on their 'comfort level' with different activities in a course (Wieman, 2015). For the study, an instructor asked students if they felt iClicker® technology would be effective in the course. The students were not familiar with the technology and 80% stated they did not want to use the technology (Weiman, 2015). The instructor chose to use the technology in the course. After completing the course, the instructor asked if the students felt the clicker technology was effective in learning new concepts and 80% responded they felt it was helpful in clarifying and understanding new concepts. Students may not have a basis to provide informed feedback; they may provide opinions not rooted in their learning, experience, or best practices. Relative to teaching excellence and assessing faculty performance, this can have profound effects on faculty members' view of their teaching effectiveness.

While students observe various factors in each class, measuring impact about teaching effectiveness via student course surveys remains challenging. Students may not always be in a position to assess the faculty members' effectiveness. For example, a faculty member needs to be enthusiastic, but not too enthusiastic. How do students perceive too much enthusiasm? On a scale of 1-5 (with 5 being highly enthusiastic), do students learn best if the faculty member is a 5? Alternatively, is too much enthusiasm more like a 3? Students may have different perceptions of how much enthusiasm they want. Other factors faculty manage which students influence include class preparedness, current knowledge levels upon entering the class, and motivation. Through

the years, faculty navigate these factors as they attempt to achieve students' learning outcomes in a course.

Schein (2013) introduced his concept of "humble inquiry" (p. 2) for communicating and asking questions of individuals. When asking students about what makes teachers excellent, there can be inherent statements made that faculty members' performance is or is not excellent. After instructors have prepared and delivered a course, students offering feedback to improve faculty members' performance – whether excellent or not – should utilize the expertise of teachers' roles and needs to suggest specific actions. Faculty members' performance in teaching is personal. Many are well versed and experienced in the topics on the courses they are teaching and want to see students learn and potentially gain the enthusiasm they have for their field. Therefore, if we are going to strive to be excellent teachers and to ensure student learning outcomes occur – everyone should have an opportunity for feedback on what is and is not working in their courses.

Finally, students frequently receive requests to complete surveys. For example, students complete end of the term evaluations and occasionally mid-semester evaluations. There are faults with these evaluative activities. Often times the students hurry to complete SETs at the end of the term (Frost & Teodorescu, 2001). Are students in the best position to answer questions about their learning, pedagogical techniques, and current events on the course? Often the faculty member is asked to leave the room to ensure student feedback is confidential. If students need clarification on a SET question, there may not be anyone in the room to address it (Cashin, 1999). Thus, the question remains, are students in a position to evaluate teaching and learning and teaching effectiveness in the SETs.

METHODOLOGICAL EXPLANATION

Both qualitative and quantitative data collection methods assisted with understanding the role of teaching excellence during the TEP. Utilizing exploratory sequential methods design as defined by Creswell and Plano-Clark (2011), the TEP first utilized a qualitative design followed by a quantitative design. With this chosen design, qualitative, face-to-face interviews informed the development of an instrument to gain input from faculty and administrators identifying teaching effectiveness criteria, which contribute to successful student learning outcomes.

The TEP required a pragmatic approach to understanding teaching effectiveness from the perspective of current faculty members, administrators, and eventually students. Qualitative and quantitative methods collected and merged data from each of the phases of the project building on the previous phase. Having the opportunity to summarize the responses from the qualitative interviews allowed the exploration of consistencies and differences in the feedback of the faculty members and administrators. The constructivist paradigm allowed for the integration of additional information from other universities and best practices found in the literature to inform the development of a quantitative instrument to assess teaching effectiveness. The quantitative instrument also incorporated the details from the interview responses faculty and administrators provided. The overarching purpose of the TEP eyed an opportunity for the possibility of an implemented solution by faculty members each term.

The development process of three instruments evolved with input from the TEP advisory committee members, faculty members in the department. Aligned on the goal to revise the departmental code with excellent teaching definitions and assessment methods, the TEP members agreed on the need to align on a definition of excellent teaching for faculty members. This definition would better incorporate student success and learning in courses.

Individuals deemed important for the project provided much data and input. The advisory committee was reluctant to ignore the quantitative data available through the student evaluation of teaching (SET) records and recognized they played one role in assessing teaching effectiveness for the project. The SET records provided the perspective of students on what was, was not working in the class, and was a single input record. By starting with the face-to-face interviews and then analyzing the course evaluation data, the difference and/or associational aspects of teaching effectiveness from the faculty member and students' perspectives occurred.

The student evaluation data available from 10 fall and spring semesters (a span of four years) allowed the use of the post positivist paradigm for assessment. We attempted to interpret the interview responses from faculty and administrators and verify if there was alignment with the students' course evaluation data. The TEP members recognized each of these paradigms' contribution to the successful completion of the project.

INSTRUMENTS AND SAMPLES

Throughout the project, data and lessons learned informed the next phase. For example, as part of the exploratory sequential portion of the project interviews allowed participants to "share their expertise with an interested and sympathetic listener" (Merriam & Tisdell, 2009, p. 107). Their expertise and input evolved to eight themes representing teaching effectiveness. While not utilizing qualitative analytic software, the advisory committee agreed on the themes from the interviews.

There were three instruments developed for this study aligned with the overarching tasks of the TEP. The first instrument was the interview instrument to capture the voice of faculty members' and administrators. The second instrument used best practices to identify criteria necessary for effective teaching and achieving learning outcomes. The final instrument captured teaching effectiveness from the perspective of the students. The outcome of these instruments

provided clearly articulated expectations of teaching for faculty with higher teaching workloads

assigned. Thus, the TEP successfully achieved the four tasks put forth by the department.

Instrument Development to Capture the Input of Key Personnel

At the first advisory meeting, discussion focused on the need to review available

educational research on teaching success criteria. This task focused on learning the benefits and

challenges of defining higher education performance. There was the need to solicit departmental

faculty members, administrators, and pertinent other faculty members for their input and

perspectives on the topic of teaching and learning effectiveness. The advisory committee

brainstormed questions to capture input in the interviews on what faculty members and

administrators deemed necessary for teaching effectiveness. The less structured brainstorming

relied on the advisory committee members' views on what was important for teaching

effectiveness. During this stage, research on teaching effectiveness continued. The additional

research and brainstormed questions dictated revisions of the interview questions with questions

added to the instrument for faculty and administrator interviews.

Finalizing the questions for the instrument required several iterations. Each member of

the committee provided input to the wording of the questions and what questions to ask. The

advisory committee members each had an area they felt should be included in the interviews. In

retrospect, an outline of the critical areas to focus on relative to the topic of teaching

effectiveness could have streamlined the process of developing the interview prioritization

instrument. Because of my lack of experience in higher education teaching, the process required

several meetings to reach consensus on the final interview instrument.

The pilot study consisted of an interview with one identified administrator. It was

apparent during the interview and further review that the instrument was weak in a number of the

questions. The purpose of the pilot study was to vet unclear questions and allow the opportunity

to make the instrument stronger. By utilizing the years of experience from the advisory committee and educational research, the next version was much stronger and clearer. I relied on the experience of the faculty members on this process and realized that I would have to take a stronger stance in the process if we were going to be able to successfully define and assess teaching effectiveness. For example, the first question was "What are your thoughts on a quality initiative to measure teaching effectiveness?" The response in the pilot study indicated the question was too vague. The feedback suggested defining teaching effectiveness in the department first and then addressing the remaining teaching effectiveness questions.

The collaborative nature of the project, with the involvement of the advisory committee, required clarification for several statements and multiple revisions to the instrument before proceeding to the rest of the interviews. For the final draft of the instrument, seven new, more focused questions replaced six of the original items. For example, the advisory committee replaced the noted weaker question with "What excites you about this quality initiative to measure teaching effectiveness in [this department]?" The advisory committee reviewed the revised instrument prior to scheduling additional interviews. While clarifying and changing some of the wording, the committee agreed the questions addressed the areas identified as important and agreed it was a more robust instrument (Appendix B) than the original instrument to assess teaching effectiveness.

Faculty members interested in having their voices heard volunteered for interviews after the department head communicated that option at the beginning of the project. The GRA coordinated with faculty and administrators and conducted interviews during the fall 2014 semester (as detailed in Table 3).

The interviews ($n = 26$) produced responses to initiate changes to the instrument. During the remaining 26 interviews, we received more detailed responses: "value of continuous improvement is interesting to faculty", "the environment for learning is influx", and "it [the department] is getting to be a richer teaching environment".

Instrument Development – Identify Best Practices

Collection of the views and voices of the faculty and administrators on teaching excellence finished. Themes from the interviews coupled with reviewed research literature and discussions with the advisory committee provided insights to what faculty members valued. At this point, the focus shifted from faculty members' perception of excellent teaching and learning to identify best practices used by excellent teachers. The definition of 'excellent' teachers were those who helped students succeed and learn, which incorporated their desire to identify individual students in their class and to offer opportunities for all students to learn.

From this definition, the goal was to provide faculty with a reasonable number of best practices, which defined excellence from the input of themselves, their peers, administrators, and educational research. These evolved from the qualitative interviews, were measurable, and were achievable in courses each term by faculty members.

Table 3. Interviewee Departments, Position, and Completion Schedule.

Interviewees	Number	Month				
		August	September	October	November	December
Internal Department						
Professor	5	2		2		1
Associate Professor	3		1	2		
Assistant Professor	4	1	1	1	1	
Adjunct Professor	3		1	1		1
Administration	2		1	1		
Other Departments						
Administration	6		3	3		
Professor	2			2		
Faculty	1		1			
Total	26	3	8	12	1	2

The last column of the Department Constructs Developed table (Table 4) represents the final eight constructs, identified from the interviews and research, with feedback from the advisory committee listed as original and interim constructs.

This subset of best practices required the identification of specific practices, which exemplified success. Numerous iterations (17 to be exact) incorporated the input of the advisory committee members. It did not lack input of opinions of what should be done next for the project, yet many suggestions were perceptions, not rooted in educational research. Given the newness of the topic and experiences of the advisory committee, we attempted numerous times to incorporate feedback and suggestions. Attempts to gain consensus were unsuccessful. At this

point in the project, it was becoming clearer that differing perspectives of the members of the advisory committee was challenging the advancement of the project. Division on what effective teaching was and which assessment strategy to utilize was yet to be determined. The educational research anchored the remaining revisions in the instrument containing the best practices.

Table 4. Departmental Best Practices and Constructs Developed.

Original constructs	Interim constructs	Final Constructs
Academic challenge	Academic challenge	Academic challenge
Appreciation of the subject	Appreciation of the subject	Student appreciation of the subject
Assessment/evaluation of students	Best practices/course design	Assessment of students
Community	Facilitate community	Community
Course content	Organization of course content and pedagogy	Course organization and pedagogy
Critical thinking	Student learning/critical thinking	Opportunity for student learning
Student engagement	Instructional delivery	Student engagement
Interaction	Interaction	Student/instructor interaction

An interim step in the instrument development process was to align the selected constructs with the defining criteria chosen. Leveraging the work done at other colleges and universities, such as The Ohio State University, Rose Hulman Institute, Eastern Illinois, Marshall University, and Colorado State University TILT Learning Ecologies Assessment to name a few, guided the work with the constructs and specific criteria alignment.

The next instrument compiled the constructs (8) and specific criteria (51) identified from the interviews and narrowed them to provide a more realistic number of criteria to focus on when teaching. For example, to represent academic challenge, one of the specific practices was 'course activity reflected the appropriate effort for the course level (i.e., 200-level, 400-level).' Another example for course organization and pedagogy was 'statement of course learning objectives in

the syllabus.' With the various iterations and construct titles identified, the advisory committee had achieved clear and concise wording of the best practices. Much of the work done between meetings to gain full agreement among the members was challenging. Departmental faculty members' schedules, between teaching, research, and student meetings slowed the progress of the project. While bi-weekly advisory committee meetings were scheduled, full attendance was lacking. Once meeting minutes were distributed, there would consistently be lengthy responses with differing perspectives shared. We would put the item on the next advisory committee meeting agenda for additional discussion, which added to the longer timeframe of the project. To progress, multiple meetings and discussions occurred between the bi-weekly advisory meetings. Attempting to get consensus from faculty members was challenging. Capturing the perspective of each advisory committee member was progress on the project and important for departmental buy-in. When meeting with the next advisory committee member and sharing the new perspective, the meeting would inherently end in a contradiction and/or a different perspective. At this point in the project, during the bi-weekly meetings, slow progress and consensus existed.

The final iteration of the constructs selected were: academic challenge, student appreciation of the subject, assessment of students, community, course organization and pedagogy, opportunity for student learning, student engagement, and student/instructor interaction with 51 criteria detailed to identify success. Next, we asked the departmental faculty to identify approximately 30% of the criteria for each construct, which they felt best represented effective teaching practices (Appendix C). Identifying 30% was a reasonable distribution of practices for faculty members to be able to select from in any given term. The 30% factor for statements represented approximately 19 statements faculty members could incorporate into their classes to ensure successful student learning outcomes. From the above, with responses to the

instrument from the advisory committee, faculty, and administrators, 22 statements illustrating teaching excellence were identified to represent best practices.

There were still questions as to who should assess teaching performance? Knowing there was value in input from students, the advisory committee looked at next steps while reviewing the research on students' roles in assessing courses. What are students evaluating when they provide feedback? Are they evaluating their 'satisfaction' in a class (Stark & Freishtat, 2014)? Do they understand the best practices they respond to (Layne, 2012)? Are students able to evaluate their learning (Weinberg, Hashimoto, & Fleisher, 2009)? When do they know what they have learned? Is a lag in time and experiences required to realize their optimal learning? These questions, several identified in the literature, and others discussed at advisory committee meetings, informed the next step of the study, which was the development of a student instrument to assess teaching effectiveness.

Student Instrument Development Process

To achieve one of the TEPs goals, questions needed to be targeted to address the students' role in providing input to teaching excellence and student learning outcomes. The 22 statements faculty members selected in the last phase represented practices excellent teachers utilize to ensure student success and learning outcomes. Yet, assessment of faculty members' performance from students' perspective was missing. The gap identified required a multi-phased design. The multi-phased design was building on the learnings from the previous phases of the research (Creswell & Plano-Clark, 2011). What should determine if students were in a position to address faculty performance and with which questions?

The advisory committee began discussions about the value of students responding to the statements as a secondary feedback option. Research had shown often students rate faculty members' performance on a SET, but they may not be capable of accurately assessing

performance (Layne, 2012). Braga et al. (2014) found students' evaluation of overall teaching

quality related to the effort they expend in a course, their realized utility of the course, and what

they observed in the classroom more than the actual quality of the teaching. The advisory

committee wanted input from students to evaluate teaching effectiveness. Students could play an

integral part in providing input to evaluating teaching if questions addressed what they are

capable of assessing. The advisory committee members agreed on the shortfalls of the university

issued current SET instrument and invested time and energy to develop an alternative assessment

tool.

Students could provide feedback on the items faculty members identified as important to

successful learning outcomes. Much work went into developing criteria for students to rate

teaching effectiveness based on a 7-point Likert scale of Strongly Disagree (1) to Strongly Agree

(7) with the newly developed SET instrument. Faculty and advisory committee members

completed over 12 iterations of the instrument.

STUDENT INSTRUMENT PILOT TEST

The advisory committee determined the best practices for student input on a course

survey, yet the student-based wording of the criteria hoped to ensure valid and reliable responses.

The goal was a less rushed and more valid student-based instrument distributed and received

from students earlier in the term. The consensus was to have a student survey distributed in the

14 or 15th week of the 16-week semester. An aggressive target was set to develop and pilot test

an instrument within three weeks during the spring 2015 semester. Feedback from this pilot

study would inform improvements for future iterations of the instrument. With wording directed

at students, three classes agreed to pilot the new instrument. Appendix D details the questions

utilized in the study. Faculty members volunteered 15 minutes of class time in the 13th and 14th

week of the semester to have students complete the survey focusing on the wording of practices,

not their responses to the practices. The pilot study was performed in two advisory members' classes and one faculty member's class within the department by the research assistant. Responses were from 22, 26, and 72 students ($n = 120$). With one response incomplete and unusable, 119 usable responses were enough to drive future improvements on the instrument.

This part of the project was working toward a valid and reliable instrument completed by students in each course, each semester, in the department. This instrument would supplement the university issued SETs. This instrument provided constructive feedback to faculty members on what was working in the course and what to improve from students' perspectives. The advisory committee anticipated the need for changes in the intended practices from the pilot test. It was important to focus on the validity of the data and responses to ensure rigor in the actionable feedback for faculty members' performance each term. Without valid information, there was both the potential for students to provide inaccurate input; as well as, for faculty to focus unduly on modifying portions of classes for change sake. As work done by Layne (2012), Stark and Freishtat (2014), and Weinberg et al. (2009) shows, it is important to provide clear and concise statements for students responses relative to teaching and learning excellence. The wording of each of the practices on the student feedback instrument will need revisited periodically to ensure they remain relevant to current best practices. The robustness of this part of the analysis was necessary to revise the instrument and provide constructive feedback for faculty members.

The students' feedback instrument continues as a work in progress. Next steps needed to complete the students' feedback instrument would be to assess the data from the pilot study, adjust questions as necessary, and then pilot the revised instrument in classes to assess the validity and reliability of the instrument.

DISCUSSION

Projects initiated to assess faculty members' teaching need to take into account the culture of the organization and ensure the practices are clearly articulated. The TEP was no different. An assumption going into the project was faculty and administrators supported the project, with assessment criteria needed for the newly defined teaching loads. However, there was minimal alignment among departmental faculty members on teaching assessment for both tenure and promotion and as a member of the team. When increasing the teaching load from 50% to 70%, some faculty and administrators felt the increase in teaching would directly correlate to a decrease in research expectations. Yet others believed the research requirements should remain the same because some research could (or should) be drawn from or aligned with teaching and learning. It was difficult to find consensus on the role teaching played in the department relative to student learning outcomes and tenure and promotion. The evolution of the project highlighted how the department viewed the new workloads for teaching and learning excellence. Most faculty members agreed that all faculty should strive for excellent teaching and successful students' learning outcomes.

This project focused on providing additional indicators for defining and assessing teaching performance and successful student learning outcomes for the department. The newly developed constructs, best practices, and student instrument could provide input on the faculty members' teaching effectiveness. It would be remiss not to mention other, generally approved sources for assessing teaching performance and learning outcomes. Students' Evaluation of Teaching (SET) instruments typically provided at the end of a term are one of different inputs to assess teaching performance. Faculty members may utilize the university issued and newly created SET feedback forms, as they deem fit, to improve performance and/or to continue activities in a course based on the students' feedback. As discussed in the manuscript, there are

58

other effective methods of assessing performance such as peer review, course material review, and mid-semester evaluations.

Additional evaluation methods discussed as part of a task force initiated in 2011 by The Institute of Teaching and Learning (TILT) at Colorado State University provided insight on faculty performance. Faculty who place a high importance on their skills as teachers, including those recognized via distinguished professor awards, voluntarily joined the task force to help develop resources for their peers. These resources targeted addressing the importance of teaching and course outcomes. The underlying premise of the task force was that excellent teaching and learning are necessary for students to excel in and after a course. The 2011 recommendations focused on assessment (#1-3) and professional development (#4) items for teaching effectiveness:

1. Teaching effectiveness should be assessed in part through the use of teaching portfolios during merit, promotion, tenure, and post tenure reviews. The University should develop a web-based portfolio system that will allow faculty members to provide evidence of teaching effectiveness.
2. Teaching effectiveness should also be assessed through peer-observation of teaching.
3. Assessments of teaching effectiveness should include the faculty member's reflective statements on teaching performance and activities.
4. Existing professional development programs supporting teaching effectiveness – in TILT, in the colleges and departments, and in student affairs – should be continued or enhanced. These groups should collaborate on the development of new professional development programs supporting teaching effectiveness (TILT, 2011, para 1).

While these were the recommendations of one task force, educational literature contain similar suggestions. There is much research on similar task forces and projects striving to define and assess teaching and learning performance. The TEP wanted to keep students and their learning at the forefront of the discussions focusing on recommendations #1 - #3 above. By focusing on teaching performance, we may, or may not, miss the critical outcomes – student success.

The job description and objectives of the TEP focused on four areas: 1) evaluate and assess teaching performance for award of tenure, 2) provide voice of faculty, 3) identify best practices from teaching excellence research, and 4) define expectations of teaching with higher teaching workload faculty. Within the case study, best practices from teaching excellence research and the voice of faculty informed the development of expectations of teaching for assigned higher teaching loads. These expectations evaluated and assessed teacher performance. While the final recommendations were not completed nor added to the performance code, the advisory committee reviewed a first draft of the updated performance code.

IMPLICATIONS FOR OTHER HIGHER EDUCATION DEPARTMENTS

Any teaching and learning excellence project should begin with a review of the higher education codes at their institutions. These collaborative codes describe the cultures of the university, college, and departments tasked with teaching effectiveness and may influence defining and assessing teaching and learning excellence. The goal of this case study was to utilize educational research and the input of faculty members and students to define and assess teaching effectiveness in one university department.

There are several suggestions to ensure a successful teaching effectiveness project after reviewing performance codes. First, it is highly recommended to utilize cross-functional and varied hierarchical committees to capture the voice of faculty and administrators. To ensure successful student learning outcomes, the buy-in of faculty and administrators is crucial. Second, the objectives of the project articulated from the administrators and communicated to faculty members will ensure a smoother execution of the project. Otherwise, consensus must be achieved on the objectives of the project or risk a slowed project. Third, project leadership to ensure communication among administrators, faculty members, and the project team can help alleviate any frustrations from the groups participating. Communication is important to reinforce

buy-in from faculty members. Finally, acknowledgement of an outsider coordinating the project, while requiring more work for the outsider, does add an additional perspective to teaching effectiveness and the culture of the department. Fresh questions can lend insight on what is or is not working well.

One size does not fit all when defining and assessing teaching excellence. What one teacher does in a course – deemed successful, may not work in another course. Course content varies and faculty members should be encouraged to focus on practices that have worked for successful students' learning outcomes. Thus, it is important to have resources with various opportunities for faculty to excel. Based on the work of the TEP, the various inputs to assessing teaching excellence include mid-semester evaluations, faculty portfolios, and peer evaluations. Table 2 – Examples of Evidence of Effective Teaching shows additional subjective methods of teaching effectiveness. A department must identify and support best practices in their discipline(s) for teaching effectiveness.

SETs represent students' perceptions at a singular point in time. During a course, faculty members balance numerous responsibilities: assignments, exams, and projects. Each of these activities assume the assessment of students' learning. Faculty members attempt to ensure learning is consistent with the learning objectives for the appropriate level of the course. Research has shown a direct relationship between the SET responses and students' performance on the assignments, exams, and projects (Beleche et al., 2012; Braga et al., 2014; Langbein, 2008; Weinberg et al., 2009).

There were numerous learnings from the teaching effectiveness project. Consensus is a good way to get sustainable agreement though it slowed the progress of the project more than anticipated. Ensuring the advisory committee members aligned with the scope and activities of

the TEP allowed them to help articulate the work to their peers. Further, acknowledging each member of the advisory committees' reasons for joining and utilizing consensus allowed the TEP to work through the steps of the project together.

In hindsight, there are a few things, which could have allowed the project to progress more effectively. First, while bi-weekly meetings strove to keep the project moving forward, getting consensus was time-consuming. Instead of trying so hard to get consensus, I could have better articulated the supporting educational research for each of the suggestions presented. Possibly synthesizing the research would have made it easier for TEP members to stay abreast on current recommendations. Second, administrative expectations changed the direction of the project during the academic year from focusing on teaching excellence for higher teaching load faculty to teaching for all members in the department. It changed the dynamic of buy-in from faculty for new faculty members to buy-in on the purpose of items, which could influence their own performance assessment. It was important for an advisory committee to have a lead representative to manage the changes and to help get the support of the advisory committee members. With my lack of higher education experience, I struggled navigating the changed direction. We lost time trying to navigate the change with the development of multiple instruments, which would now apply to all faculty members in the department instead of the higher workload faculty members. Inviting the administrators to the bi-weekly meetings to reset expectations for the advisory committee still may have slowed progress, though it may have slowed it less. When the project ended, the advisory committee had made as much progress as we were able with such a large scoped project. We did not achieve all of the original goals of the project though the work done provided guidance to the department.

When Boyer (1990) placed teaching on the same level as discovery, integration, and application, he furthered the conversation for the importance of teaching. Today, there are 2,618 accredited, four-year colleges and universities according to the Association of American Colleges & Universities (AAC&U), (AAC&U, 2010). One way for students to learn is for the transfer of knowledge to occur between the faculty members and students. Successful learning outcomes are critical to the success of the students in the course, the next course, their careers, and their next jobs! Only when faculty and students engage during a course can success be possible – for both the teacher and the student! We must continue striving for those success factors! By attempting to define and assess teaching excellence, we moved one-step closer to teaching effectiveness and student learning success.

REFERENCES

American Association of University Professors. (2010). Tenure and teaching-intensive
        appointments. Retrieved from http://www.aaup.org/report/tenure-and-teaching-intensive-
        appointments.

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student
        learning? Evidence from an objectively graded post-test. *Economics of Education
        Review*, 31, 709-719.

Boyer, E.L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate.* New York, NY:
        The Carnegie Foundation for the Advancement of Teaching.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of
        professors. *Economics of Education Review, 41*, 71-88.

Carnegie Foundation for the Advancement of Teaching. (1998). *Reinventing undergraduate
        education: A blueprint for America's research universities.* Princeton, NJ: Boyer
        Commission on Educating Undergraduates in the research university.

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

Colorado State University Academic Faculty and Administrative Professional Manual (2013).
        Retrieved from http://facultycouncil.colostate.edu/faculty-manual-section-e/#E.12.1.

Colorado State University [department] college code (n.d.). Retrieved from
        provost.colostate.edu/codes/

Colorado State University – The Institute for Learning and Teaching (TILT) (2011). Evaluating
        teaching effectiveness task force recommendations. Retrieved from
        http://tilt.colostate.edu/sotl/taskForces/assessingTeaching/.

Creswell, J.W. & Plano-Clark, V.L. (2011). *Designing and Conducting Mixed Methods
        Research.* Thousand Oaks, CA: Sage Publications, Inc.

Czikszentmihalyi, M. (1982). Intrinsic motivation and effective teaching: A flow analysis. *New
        Directions for Teaching and Learning: Motivating Professors to Teach Effectively, 10*,
        15-25.

Dezure, D. (1999). Evaluating teaching through peer classroom observation. In P. Seldin (Ed.),
        Changing practices in evaluating teaching: A practical guide to improved faculty

performance and promotion/tenure decisions (pp. 70-96). Bolton, MA: Anker Publishing Company, Inc.

Frost, S.H. & Teodorescu, D. (2001). Teaching excellence: How faculty guided change at a research university. *The Review of Higher Education, 24*(4), 397-415.

Grassian, D. (2013). Constructive ambiguities: The inspiring, deflating, transformative, and limited possibilities for assessment in higher education. *Journal of Excellence in College Teaching, 24*(2), 155-173.

Higgerson, M.L. (1999). Building a climate conducive to effective teaching evaluation. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 194-212). Bolton, MA: Anker Publishing Company, Inc.

Huber, M.T. (2002). Faculty evaluation and the development of academic careers. *New Directions for Institutional Research, 114,* 73-83.

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

Lattuca, L.R. & Domagal-Goldman, J.M. (2007). Using qualitative methods to assess teaching effectiveness. *New Directions for Institutional Research, 136,* 81-93.

Layne, L. (2012). Defining effective teaching. *Journal of Excellence in College Teaching, 23(*1), 43-68.

MacNeil, L., Driscoll, A., & Hunt, A.N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40,* 296-303. Published online: 05 December 2014. DOI: 10.1007/s10755-014-9313-4.

Mann, K.V. (2010). Self-assessment: The complex process of determining "How we are doing" – A perspective from medical education. *Academy of Management Learning & Education, 9*(2), 305-313.

Marincovich, M. (1999). Using student feedback to improve teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 45-67). Bolton, MA: Anker Publishing Company, Inc.

Merriam, S.B. & Tisdell, E.J. (2009). *Qualitative Research: A Guide to Design and Implementation.* (4th ed.). San Francisco, CA: Jossey-Bass.

Miller, J.E. & Seldin, P. (2016). Changing practices in faculty evaluation. *American Association of University Professors*. Retrieved from: http://aaup.org/.

O'Neill, G.P. (1988). Teaching effectiveness: A review of the research. *Canadian Journal of Education, 13*(1), 162-185.

Palermo, J. (2013). Linking student evaluations to institutional goals: A change story. *Assessment & Evaluation in Higher Education, 38*(2), 211-223.

Palmquist, M. (2011, December). *University Distinguished Teaching Scholars - The Institute for Learning and Teaching Assessing Teaching Effectiveness Task Force Recommendations.* Retrieved from http://tilt.colostate.edu/sotl/taskforces/teaching/recommendations.pdf.

Porter, S.R., Rumann, C., & Pontius J. (2011). The validity of student engagement survey questions: Can we accurately measure academic challenge? *New Directions for Institutional Research, 150,* 87-98.

Rice, R.E. (2002). Beyond *Scholarship Reconsidered:* Toward an enlarged vision of the scholarly work of faculty members. *New Directions for Teaching and Learning, 90,* 7-17.

Schein, E. (2013). *Humble inquiry: The gentle art of asking instead of telling.* San Francisco, CA: Berrett-Koehler publishers.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83(*4), 598-642.

Stark, P.B. & Freishtat, R. (2014). An evaluation of course evaluations. University of California Berkeley. Retrieved from: http://senate.csusb.edu/Reports/evaluations_paper.pdf. 1-15.

Weiman, C. (2015). A better way to evaluate undergraduate teaching. *Change.* January/February, 6-15.

Weinberg, B.A., Hashimoto, M. & Fleisher, B.M. (2009). Evaluating teaching in higher education. *Journal of Economic Education,* Summer, 227-261.

Williams, K.M. (2015). *Doing Research to Improve Teaching and Learning: A Guide for College and University Faculty.* New York, NY: Routledge.

STUDENTS' EVALUATION OF TEACHING: ONE EXPERIENCE – HOW WILL YOU USE THEM?

SUMMARY

Defining and assessing teaching effectiveness is a topic covered frequently in educational research. Student course surveys attempt to provide students' feedback on faculty performance and contain both formative and summative information. Using 15,858 course surveys, this article analyzes students' feedback for rating the course and their instructor. Students rated courses above average or excellent 78.2% of the time. Yet, they rated their instructors higher with above average or excellent responses 86.8% of the time. Male instructors received lower means on both questions, respectively and rated the course 4.15 (out of 5.00) and the instructor 4.47. When assessing the level of the course, the results were statistically significant with students in higher-level courses rating the course and instructor higher. The data found in this study provide insights into some of the bias and trends in student course surveys, which faculty members and administrators could and should be aware of when using course evaluations for formative and summative evaluations.

INTRODUCTION

Teaching is a highly personal task. With faculty assessment often relying on students' feedback of the "teacher's characteristics and teaching" (Pan et al., 2009, p. 74), how do faculty members focus their attention on students' interest and the materials, which need to be communicated? Frost and Teodorescu (2001) ascertain "teaching happens in all interactions between faculty and students" (p. 410). Techniques and methods used in a classroom reflect faculty members' personal beliefs for the best ways to convey information to achieve learning outcomes (Grassian, 2013; Layne, 2012). Czikszentmihalyi (1982) posited "the real task of a professor is to enable the learner to enjoy learning" (p. 18). His work in intrinsic motivation

highlights the outcome of teaching when students want to gain more information on the topic. Yet, numerous factors can influence students' desires to learn – the faculty member is one factor. Therefore, what do effective faculty members do during a course to ensure student learning? How can activities influence students' learning and the desired course outcomes? What role should SETs play with students' responses to teaching effectiveness?

In the Boyer report, the scholarship of teaching in higher education is an important focus for faculty, along with the scholarship of discovery, integration, and application (Boyer, 1990). Boyer asserted the need to focus on the scholarship of teaching with "great teachers creating a common ground of intellectual commitment. They stimulate active, not passive learning and encourage students to be critical, creative thinkers, with the capacity to go on learning after their college days are over" (Boyer, 1990, p. 24). Bain (2004) argued the need for teachers to recognize the complexity of "human learning" (p. 8). There is one way to stimulate learning and it requires the constant evolution of teaching. Acknowledging the differences of course materials, student learning styles, and faculty dynamics in delivering the course helps to ensure better student outcomes. While student outcomes are a factor for assessing faculty performance, it is important for faculty to understand all the criteria for assessing their level of effectiveness.

Teaching requires faculty members to establish "clear goals, adequate preparation, appropriate methods, significant results, effective presentation, and reflective critique" (Huber, 2002, p. 75) for each course. The course materials (Huber, 2002; Layne, 2012), delivery methods, and different pedagogical techniques for courses taught are chosen by faculty members.

There are factors faculty have less influence over when teaching a class: their appearance, enthusiasm, energy level (Czikszentmihalyi, 1982; MacNeil, Driscoll, & Hunt, 2014; O'Neill, 1988) or those influenced by the organization or administrators -- time class scheduled,

technological resources (and their reliability)**,** and physical environment. Corrigan (2012) identified a list of 15 items administrators could or should do to support faculty members in promoting learning. The items revolve around the scholarship of teaching and learning and ensuring administrators stay current on the research.

While students observe these factors faculty have less influence over in each class, these criteria are difficult to define and assess. For example, a faculty member needs to be enthusiastic, but not too enthusiastic. How do students perceive too much enthusiasm? On a scale of 1-5 (with 5 being highly enthusiastic) do students perceive they learn best if the faculty member rates a 5? Alternatively, is much enthusiasm more like a 3 rating? Stark and Freishtat (2014) challenged the ethics of the Likert scales and whether the difference between a 1 and 2 rating is the same as the difference between a 4 and 5 rating? When averaging the SET ratings to assess faculty performance, the Likert scale ratings difference between a 1 and 2 rating or a 4 and 5 rating influences the results. If faculty members and administrators are using these scores to gauge faculty members' teaching effectiveness, they must align on the design and interpretation of the scale.

Students and parents represent additional factors influencing faculty performance through their continued active roles in setting expectations of the education experience and success following graduation (Carnegie Foundation for the Advancement of Teaching, 1998). Student and parent expectations of what is taught, how it is taught, and with what resources it is taught no longer follow the "traditional modes" of teaching (Carnegie Foundation for the Advancement of Teaching, 1998, p. 1); instead the expectations are created based on students, students' families, employers, and graduate school opinions to name a few. While there are accreditation standards, outside influences affect faculty members' performance. The focus on costs and performance

influences teaching effectiveness and evaluations. Further, how the information is used to improve learning outcomes continues to evolve and remains a popular discussion in educational literature. With new, evolving opportunities in higher education, conversations with various stakeholders needs to occur. From new and emerging degree programs to the mode of delivery of education, students and faculty, as well as, administrators' expectations should align. The students' success may be due to the faculty members' style though various factors can influence the students' performance. There is alignment identified within the research that assessing faculty member roles as teachers, guides, coaches, and collaborators relative to student success criteria should happen (Czikszentmihalyi, 1982). The dissension for the role of SETs are not if it needs to be an input to assessing teaching effectiveness, but how large of a role it should play.

Teaching is further under scrutiny due to high student debt, graduation rates reflected as the number of years needed to complete a degree program, and success in the workplace, with faculty performance used as a contributor of success. The changes within the higher education system influences faculty, students, and administrators. As an increased pressure on performance assessment for faculty, Miller and Seldin (2016) highlighted the most common method used to evaluate faculty performance has been the quantitative Student Evaluation of Teaching (SET) surveys typically distributed at the end of the course. The purpose of this manuscript was to assess the constructs defined as indicators of teaching effectiveness and compare those constructs to SET data completed from spring 2011 through fall 2015 to establish their role in gauging teaching effectiveness to drive student success.

The American Association of University Professors (AAUP) (2010) published findings on the continued collapse of the faculty infrastructure with tenure track positions declining. Their report found teaching-intensive positions have "risen sharply" (p. 1) over research-intensive

positions with "the overwhelming majority of non-tenure-track appointments being teaching only or teaching intensive" (2010, p. 2). Years later the teaching intensive positions and the decline in tenure tracked positions is receiving attention in the press (Yang & Walker, 2015). Clearly articulating the expectation of the teachers is an important step to ensure stakeholders' perspectives align on teaching excellence. Ensuring there is a way to assess these articulated expectations should be a focus in higher education for the success of faculty members. The availability of SET scores, for multiple stakeholders, such as students, parents, and administrators, should not be the primary input of evaluation for faculty performance; instead, SETs should represent one form of assessment with additional recommendations provided to drive student success.

An area of contention for assessing teaching effectiveness is the use of SET data with no standardized benchmarks. Typically, SET mean scores for each question appear on evaluation forms. The mean score may not clearly represent what the question is trying to assess. The amount of variation represented in the responses (Beleche, Fairris, & Marks, 2012) and other factors are absent with a reported mean score; thus, reporting standard deviation scores for SET questions would add more clarity to the assessment. Examples of other factors influencing mean scores are the absence of the scores of students who started the class but did not complete it or students absent when SETs were completed. Further, mean scores ignore other influences such as the level of the course, lower division or upper division, the grades students expect to receive (Langbein, 2008), or the gender of the faculty member (MacNeil et al., 2014). With several factors influencing students' responses to the SETs, it is recommended to have multiple methods to assess teaching effectiveness.

While defining how and what to evaluate, consideration should be focused on the desired

outcomes of a course. Clearly articulated expectations and frameworks to evaluate teaching

performance allow faculty members' performance to be assessed. Some faculty support the idea

that student success, as defined by the anticipated learning outcomes, should be the overarching

outcome and accountability for teaching performance (Lattuca & Domagal-Goldman, 2007). The

Lumina Foundation ("learning outcomes", 2016) defines learning outcomes with active verbs,

which students can demonstrate and use for assessment**.** Learning outcomes often are progressive

allowing the students' demonstration of mastery to attempt to gauge the faculty members'

success in delivering the materials. Yet, external characteristics associated with teaching a

course, such as students' prior knowledge and technology considerations can influence faculty

members' ability to teach effectively. Though the particulars of the criteria could vary across

departments due to differences in fields, course levels, learning objectives, or other factors,

faculty must understand the expectations of their department and institution.

Grassian (2013) argued "exceptional teaching should not be part of a learning outcome

for a course" (p. 167). He further stressed the need for the workplace to provide valuable input

on learning outcomes to ensure graduates are successful in applying their learnings. Boyer

(1990) understood effective teachers are constantly requesting feedback and by receiving

feedback, they were learning along with students and thus becoming more effective teachers

(Colorado State University, CSU Teaching Effectiveness Report, 2015, p. 21).

Lattuca and Domagal-Goldman (2007) stressed assessments of teaching should include

the faculty members' performance in conveying student learning outcomes to students. However,

as others have deemed, conveying student learning outcomes while important may not reflect

actual lessons from the class. As Williams (2015) posited, learning outcomes are "what faculty

want students to know or do as a result of the instructional experience designed" (p. 78). Faculty members make teaching decisions believed to support learning outcomes. However, to assess students' performance against the established learning outcomes during a course, course work, assignments, and exams are used to measure student learning success. While these serve as an assessment measure, are students' learning reflective in the faculty members' performance?

To ensure the achievement of learning outcomes, faculty members make decisions before, during, and after teaching a course to influence learning outcomes. For example, faculty members' first decision is typically to establish the learning outcomes expected in the course. Mann (2010) defined two types of learning outcomes, affective and cognitive in his study of self-assessment of all levels of employees in the medical field. Affective outcomes include "reactions, motivations, and self-efficacy" (p. 306). So, what role do SETs play in assessing students' learning outcomes? In addition, are there better indicators of faculty performance influencing student success? Yet, SETs often ask students to indicate perceptions of satisfaction, application of new knowledge, and perceptions of their ability relative to affective outcomes. Cognitive outcomes include "understanding of task-relevant verbal information, including both factual and skill-based knowledge" (p. 306). Outcomes reflected via grades or assignments feedback are typical cognitive outcomes on SETs. These decisions are difficult for students to assess when completing the SETs, as many students are unfamiliar with how and why instructors prepare for a course.

The Boyer Commission's stance on learning outcomes met resistance when reported in 1990. The members advocated undergraduate students' learning and development while working with faculty on research projects, one-on-one allowed both students and faculty members to excel. Active learning and applying learned classroom concepts reinforces the importance of the

classroom lessons and desired learning outcomes for students. Some faculty members resisted this suggestion as they felt students would slow their research agendas. However, the students' success and learning outcomes have outweighed the resistance at some colleges and universities, such as University of Texas-Austin and Carnegie Mellon University with students' learning enhanced due to the collaborative efforts. When faculty are establishing their goals for a course and expectations for students' learning outcomes, hopefully, the Boyer Commission's suggestions are considered.

Bain (2004) in his award-winning book "*What the Best College Teachers Do*" utilizes outcomes to define excellent teaching. He identified teaching excellence "when we see evidence about remarkable feats of student learning and indications that teaching helped and encouraged those results; we learn something about developing teaching excellence when we try to discover what fostered that educational success" (p. 15). He explains that best teachers have an investment in the success of students and want to see them learn and engage in subjects. Bain recognized the challenges and complexity to defining excellent teaching and utilized actual examples of successful teachers to portray actions and expectations.

Palmquist (2011) and the TILT task force members posited "effective teaching *begins* with the recognition and application of those elements that best stimulate student learning; teaching becomes excellent through effort, through iterative adaptations of and improvements in curricular material, through honest self-reflection, through the solicitation of substantive feedback from colleagues and students, and through a spirit of humility and a willingness to continue to approach teaching creatively" (p. 6). Both Bain and Palmquist recognize more than one point in time is needed to reflect teaching effectiveness. It is a time-consuming, evolving

process for faculty members to become excellent teachers and for their students to achieve successful learning outcomes.

Palmquist (2011) and the taskforce further argue "… what students take away from a course in terms of knowledge, skills, attitudes, and abilities – are not synonymous with teaching effectiveness. Although they are closely linked, it is possible (albeit rare) to teach a course well without necessarily achieving the learning outcomes associated with course goals" (p.1). Thus, while SETs have the potential to measure teaching effectiveness, this study assessed the value SETs provided over a period for course and instructor evaluation. Further, what other indicators and assessment recommendations could provide input to better assess teaching effectiveness in higher education?

SETs typically address course and instructor factors in an attempt to measure students' perceptions of how well the faculty member taught. In addition, many of the questions are associated with items discussed thus far, which are essential to learning by the students. For example, at one university studied, the Student Course Survey, distributed at the end of each course, asked a question "How effectively did the instructor facilitate student learning?" This is difficult to assess, as Palmquist (2011) and the TILT taskforce indicated. Bain's study assessed students' qualitative responses to these types of questions to see if the comments were short term, shallow responses or longer term, 'changed my way of thinking' responses. What is the best way to measure student learning? Did the faculty member teach to the tests allowing students to achieve higher grades (Stark & Freishtat, 2014)? If there is a higher-level course the student enrolls in, does future performance in that course represent learning? Moreover, when looking at student learning, what role do faculty dynamics, time lapse between courses, and

departmental alignment of course curriculum play? It continues to be a challenge to measure if the students actually learned to the course goals.

GENDER INFLUENCE ON SETS

Factors difficult to assess are contained within the SETs. Faculty members possess "age, race, gender, experience, subject," (Langbein, 2008, p. 423) and personal characteristics. Each of these characteristics influences students' and faculty members' interaction. It is not a matter if the characteristics influence students' learning outcomes but instead how they influence students' learning outcomes. However, assessing if and how these factors influence responses on teaching effectiveness is more difficult.

Recently, MacNeil et al. (2014) reported bias related to gender in "What's in a name: Exposing gender bias in student ratings of teaching." The research study occurred in an online course assessing whether gender played a role in teaching. For their study, the researchers masked the identity of two faculty members who were teaching four classes. In two of the classes, the faculty members utilized their real names and gender, in the other classes, they taught the class the same way as when they used their real names but utilized the other instructors' name and gender. Since these classes were all online, student interactions were all online which allowed faculty members' deception to test their hypothesis of gender playing a role in students' ratings of teaching effectiveness.

The study utilized 15 close-ended questions to measure effectiveness (6), interpersonal traits (6), communication skills (2), and overall quality of the teacher (1). Twelve of the questions factored in instructors' teaching (p. 7). They found a "perceived gender identity" (p. 8) gap. The perceived male responses scored the highest of the four scores – actual female, actual male, perceived female, and perceived male. The perceived male teacher "received significantly higher scores on professionalism, promptness, fairness, respectfulness, enthusiasm, giving praise,

and student ratings index*"* (p. 8). The actual female scores were higher on skills, such as caring, enthusiasm, and praise. This disparity among the feedback for female instructors versus male instructors was intriguing. MacNeil et al. revealed female instructors are "systematically disadvantaged in academia" (p. 11). While the study would be challenging to replicate in face-to-face courses, the results of bias in SETs may be applicable. If there is gender bias in SET feedback, bias would be present in the use of these student course surveys to assess teaching effectiveness.

Boring (2016) supported MacNeil et al.'s findings that students have a consistent bias for male versus female instructors. Boring (2016) found students gave favorable ratings to male and female instructors based on stereotypes. For example, male instructors scored higher on "authoritativeness and knowledgeability, such as class leadership skills and professor's ability to contribute to students' intellectual development" (p. 28). Female instructors scored higher on "warm and nurturing" stereotypes, such as "preparation and organization of classes, quality of instructional materials, clarity of the assessment criteria, usefulness of feedback on assignments, and ability to encourage group work." (p. 28). Acknowledging biases and stereotypes by faculty members and administrators when using SETs for formative or summative assessment is critical.

COURSE LEVEL IMPACT ON SETS

Boyer (1990) posits "in teaching undergraduates, faculty confront circumstances in which more general knowledge and more precise pedagogical procedures are required." (p. 70). Yet, most students are not familiar with the fact that there are specific techniques used in courses, lower level or higher level, to ensure successful learning outcomes. Many students are not familiar with Bloom's Taxonomy (Krathwohl, 2002) and the dimensions Bloom put forward relative to student learning. While decisions on the pedagogical techniques utilized in the course are up to the faculty members, students may or may not be comfortable with the delivery of the

course materials. These techniques have the ability to influence SET responses for faculty members.

When students are providing feedback on a course, the level of the course, and how the faculty member taught the course could influence the SET responses. First, if course content was too difficult or not difficult enough, students will struggle to engage in the learning process (Porter, Rumann, & Pontius, 2011). Second, a course should advance students' prior understanding or beliefs of a concept or idea (Lattuca & Domagal-Goldman, 2007). By repeating prior knowledge, there can be benefits of the review. However, there needs to be a balance of new learning and review to ensure students have the skills and knowledge expected at the completion of the course. Finally, course activities should reflect the appropriate effort, both quality and quantity, for the level of the course and be integral to the learning objectives (Ried, 2011). Applying the dimensions of Bloom's Taxonomy (Krathwohl, 2002) and students' progressive application of information, freshmen students remember or recall new information; where seniors would be expected to synthesize gained information.

TEACHING EFFECTIVENESS PROJECT BACKGROUND

A teaching evaluation project (TEP) began in fall 2014 targeted at defining and assessing teaching effectiveness with access to 15,858 students' evaluation of course records. There was extensive data contained within the records – course number, expected grade, faculty member identification, ratings on evaluation of 29-scaled items, self-recorded attendance performance, semester, with a few additional, ancillary items. For additional comparative analysis, there was an opportunity to assess the records to identify trends relative to faculty members' performance. The TEP's purpose was to address four questions (right column Figure 3). The statements that guided the project were: Evaluate and assess teaching performance for progress toward award of tenure, provide voice to faculty in the department on the criteria used to define and assess

teaching excellence, identify best practices from teaching excellence research, and to define expectations of teaching for faculty with higher teaching workloads.

While the focus in one department at one university, teaching effectiveness research is a common theme on multiple campuses in multiple departments across the United States experiencing similar change. The current practices of academics and focus on teaching, still relevant from the Boyer Commission's recommendations, played an important role in defining and assessing teaching excellence for the TEP. Initially the SET data played a large role in the TEP with focus on identifying a student evaluation score for teaching effectiveness. It became apparent various units in the university were focusing on other methods of measuring teaching effectiveness as well. Identifying the constructs to measure teaching effectiveness and ensure student success was the priority.

BETTER UNDERSTANDING OF TEACHING EFFECTIVENESS

Faculty responsibilities are typically teaching, research, and service. Faculty often have the flexibility to choose their research topic(s), frequency of research, and research format (Langbein, 2008). More visible than activities in the classroom and sometimes easier to assess, there are more clearly defined expectations on research than teaching performance (Anderson et al., 2011; Carnegie Foundation for the Advancement of Teaching, 1998; Cashin, 1999; Rice, 2002). Because of the way faculty work independently to teach, prepare for class, grade assignments and assessments, there is the opportunity to evaluate the desired learning goals in a course or the interpretation of the topic. Comparing faculty members' performance to other faculty members (Frost & Teodorescu, 2001), such as through a peer review can be an additive option to SETs for assessment. The roles of research and service can, and should be, brought into the classroom and thus used as a goal for faculty as well.

- Clarify and improve mechanisms used to evaluate teaching performance in ways that provide constructive feedback to faculty and sustain or enhance excellence in teaching and to develop assessment methods and indicators of teaching success and, scholarship associated with teaching, which will serve as standards for award of tenure.

- Evaluate current situation to identify opportunities for new directions and potential concerns about the initiative among faculty

- Conduct a broad based assessment of innovative ideas about creating excellence in teaching (e.g., via methods of instruction, ways of organizing, reward systems).

- Delineate in both general and specific the ways to define excellence in teaching performance, teaching scholarship, and scholarship more generally for faculty on higher teaching assignments.

Evaluate and assess teaching performance for award of tenure

Provide voice of faculty

Identify best practices from teaching excellence research

Define expectations of teaching when higher teaching workloads

Figure 3. Teaching Effectiveness Project (TEP) Job Description and Objectives.

Teaching effectiveness themes developed by assessing the SET data reflected what students did say and what they did not say. The goal of the TEP was to identify departmental criteria for faculty members to understand the performance criteria expectations for their evaluation. However, the debate on the purpose of SETs continues in the literature. While various works continue to be published relative to the role of student course evaluations (Braga, Paccagnella, & Pellizzari, 2014; MacNeil et al., 2014; Smith & Gadbury-Amyor, 2014; Stark & Freishtat, 2014; Weiman, 2015; Williams, 2015), actual uses do not appear to be consistent in how the data and scores are used within the performance appraisal and/or promotion and tenure process (Brydges et al., 2012; Cashin, 1999; Dezure, 1999; Lieberman, 1999; McMillan & Hearn, 2008). For example, Cashin (1999) suggests the students' SET scores should not be an average to use when reviewing faculty performance. Averages do not take into account the range

or number of students responding. Yet, faculty members at a Research I: Doctoral Institution frequently present an average SET score per course for documenting teaching effectiveness.

Lieberman (1999) further proposed Classroom Assessment Techniques (CAT) as a way to collect formative information from students and then using that to posit instructor performance. The purpose of this study and analysis was how well would Student Evaluation of Teaching (SET) instruments, at the university studied, assess constructs deemed important for teaching effectiveness in a case study and whether the SETs should play a role in assessing teaching effectiveness.

DEFINING AND ASSESSING TEACHING EFFECTIVENESS – THE STUDENTS' VOICE

In addition to the voices of participating faculty members, the review of educational literature and other university practices identified best practices for defining and assessing teaching excellence. While peer reviews, faculty reflective statements, and teacher portfolios require the time of faculty member and peers to complete, the voices of students were missing. Student feedback was deemed an important voice. Research acknowledges that often students spend 8-16 weeks during a course with faculty members. The students' classroom experiences and the interaction with faculty members was a desired perspective to capture. These experiences helped define teaching excellence, with seven constructs identified as representing best practices. The students' voice provides an opportunity for input on the prioritized constructs listed in Table 5. There is no question on the importance of these constructs, but gauging students' perspectives of each of these constructs did come into question. We will discuss each construct created and decisions made to have students assess each construct in more detail.

Academic challenge gauges if a course was difficult enough to result in learning and not too difficult as to lose students' attention in the learning process. There must be a balance of new learning and review to ensure students gain the skills and knowledge throughout the course.

Table 5. Constructs Selected to Represent Teaching Effectiveness.

Academic challenge
Student appreciation of the subject
Assessment of students
Community
Course organization and pedagogy
Student engagement
Student/instructor interaction

Students' appreciation of a topic represented the next indicator of faculty members'

performance in a course. Students' desire to learn more about the content, recommend the course

to other students, choose to continue with advanced classes, apply the learnings to real world

issues, and their enjoyment of the topic are all indicators students are in a position to share. Do

students' responses reflect objectivity or reflect their opinion? For example, the reasons for

recommending a course to other students is difficult to ascertain – is it due to perceived easy

grading, minimal work requirements, or is it based on the level of learning incurred and/or the

relevance of the information to current events and/or interaction(s) with the faculty member? Or

does a recommendation of a course have anything to do with the instructor? The statements

selected to ascertain students' appreciation of the course content were recommendation of the

course to other students and relevance to real world issues.

Gain related to the desired learning outcomes was measured via faculty assessments or

assignments for students. Students' desire to learn more about the topic at the end of the course

and successfully achieving the stated learning objectives defined success. Multiple measures

selected for this construct were:

   i) Learning (Feden, 2012; Frost & Teodorescu, 2001; Lattuca & Domagal-Goldman,
       2007; Rice, 2002) as defined as the comparison of knowledge at the beginning and
       end of the term, linkage with course objectives, and assignments provided choices for
       students' original and creative work

ii) Feedback (McMillan & Hearn, 2008; Ried, 2011; Spooren, Brockx, & Mortelmans, 2013) with assignments returned within 7 days providing constructive feedback and multiple opportunities within the term for students to give course feedback

iii) Clearly articulated grading expectations (Drennan et al., 2011; McMillan & Hearn, 2008) with rubrics developed and provided for most assignments and assignments of grades aligned with stated criteria.

One statement selected encompassed these principles for students to review: helpful feedback was provided on assignments.

The fourth construct gauged community created in the course. Boyer (1990) stressed the importance of community as the way some faculty members consider the university, students, and peers, as family. He further posited "cultivating a sense of community" (p. 35) directly influences students with collaborative learning and interpersonal relations. While there are additional complexities in large classes, criteria utilized in creating a community environment were:

i) Group work among students (Dezure, 1999) -- opportunities for group work, groups formed with students of varying levels of competencies and abilities

ii) Trust among students in a course (Frost & Teodorescu, 2001; Seldin, 1999) -- students borrowing or lending resources to each other, interacting with each other to help better understand content, and an atmosphere of trust in the class

iii) Opportunities for improvement for the instructor (McMillan & Hearn, 2008; Ried, 2011; Spooren et al., 2013) -- course feedback periodically requested by instructor, improvements or change initiated based on feedback, and students encouraged to actively create solution(s) for issues identified in class.

Each of these statements identified by other universities attempted to gauge community within their departments with educational research literature anchoring their learnings. However, many of the statements were either too vague or too simplistic to assess attainment of community. The two statements added to the list were: I was given opportunities to contribute in class and I was comfortable contributing in class were chosen to represent community in a course.

Course organization and pedagogy represents faculty members' efforts before, during, and after a particular course. However, the actual activities and best practices for faculty members, when perceived by students can underestimate faculty members' roles and responsibilities. For example, when a class session is disrupted based on students' questions, 'delivery of class content in an organized fashion' might be under-estimated. Therefore, it was important to identify best practices, which reflect good course organization and pedagogy as listed in Table 6.

Again, while few would disagree the original items are important activities to incorporate in a course, it was difficult to measure and assess each of the items. To better articulate sustainable best practices this list was narrowed from 12 to 5 items (Column 2 of Table 6). The five statements chosen for the list to reflect course organization and pedagogy were allocation of class time for students to practice new skill(s) or technique(s), instructor used a variety of instructional methods, the instructor periodically reviewed main points and concepts, the instructor made an effort to ensure I understood course content, and instructor used real life examples.

Engagement represents students' ability to measure their own learning (Bain, 2004; Holman, 2013). It is challenging to identify what engagement looks like and the level of engagement in courses. Bain (2004) stresses the opportunities to utilize humor where appropriate to increase student engagement in the classroom. He also spoke of the ability to engage students with various technological modes, such as videos. Today, additional technological means of communication attempt to increase engagement, such as TedTalks [®], YouTube [®], and video conferences (such as, Skype [®]). With six items identified (Table 7) to describe engagement, the question still existed how to measure success on this construct. Thus, the list for students to

assess were: there were instances during the class when the course had all of my attention and I came to class prepared.

Table 6. Best Practices for Course Organization and Pedagogy.

Original Statements of Best Practices for Course Organization and Pedagogy	Final Statements of Best Practices for Course Organization and Pedagogy
Allocation of class time for students to practice new skill(s) or technique(s)	Allocation of class time for students to practice new skill(s) or technique(s)
On-time start to class sessions	
Delivery of class content in an organized fashion	
Statement of course learning objectives in the syllabus	
Use of a variety of instructional methods	Instructor used a variety of instructional methods
Instructor's enthusiasm for course content	
Summary of main points presented at the end of each class	The instructor periodically reviewed main points and concepts
Efforts of instructor to ensure student learning	The instructor made an effort to ensure I understood course content
Integration of content from other disciplines	
Logical and sequential presentation of course content	
Update of course content/materials from previous years	
Presentations of real world examples in class	Instructor used real life examples

The final construct deemed valuable for students to assess related to student and instructor interaction. Bain (2004) ties this construct in with fostering student engagement. This construct stayed as a stand-alone item due to the importance of managing the dynamics of a class. Some factors which could influence student and instructor interaction include size of class, level and type of course (undergraduate, graduate, lab, recitation), or required versus elective class. Student factors could include level of effort or attendance in the course. The expectations of student and instructor interaction would vary, both positively and negatively based on the

matrix of the above factors. Thus, the statements chosen for the list to exhibit strong performance for this construct were sufficient access to instructor outside of class, encouragement by instructor for students to answer difficult questions, wrong answers responded to constructively, and error or insufficient knowledge by instructor admitted, when applicable.

Table 7. Student Engagement Constructs and Challenges to Assessing the Constructs.

Student engagement constructs	Challenges to assessing student engagement	Final assessment questions
Students shared examples about a topic after a topic was presented	No timeline is needed for this example -- students could share information after course completed	
Classroom climate of respectful consideration for differing opinions	Level of controversy for the topic of the course could impact this construct	
Frequency of students' questions, discussion, and similar forms of participation	Size of class could impede students' comfortable level participating	
Student attendance	Various factors impact students' attendance	
Attentiveness by students in class	Various factors impact students' attentiveness	There were instances during this class when the course had all of my attention
Student preparation for class	Faculty members need to provide clear expectations of preparedness to students	I came to class prepared

After the advisory committee reviewed the responses to these best practices, some of the statements were deemed difficult to assess. For example, the competency of the instructor on a topic may deem the statement, error or insufficient knowledge admitted, unneeded. If the faculty member was competent on the questions and content of the course, there may not be a need to acknowledge an error or insufficient knowledge. The final statements selected for the list for students' assessment of student and instructor interaction were: I was given opportunities to provide feedback about the course and the instructor was responsive to helping students.

The 7 constructs and 16 items identified for students' responses at the end of a course comprised the departmental teaching effectiveness framework (Appendix C). While the intent of the newly developed instrument was to provide an additional, better assessment of teaching effectiveness, it was important to acknowledge the summative and formative role this instrument could represent. Student feedback, on these seven constructs provided one contribution for formative (continuous improvement) and summative (promotion) assessment. The voice of students represents one voice among others in the feedback process for faculty members about teaching effectiveness in a course.

While the newly developed student instrument would provide new, more data-rich feedback, the TEP still had access to over 15,000 university administered SETs. Thus, expanding on the tool development intrigued the TEP with two interest areas worth pursuing: gender and course level. These attributes may have represented potential influencers to student evaluation of teaching. With the research done on gender and student maturity, it was an area of interest to the study.

DATA

The questions used by administrators for summative, promotional decisions, and formative, teaching performance improvement, feedback are 'How do you rate this course?' and 'How do you rate this instructor?' Each of these questions listed at the end of the sections: 'about this course' and 'about this instructor' portion of the SET proxies as a cumulative score of faculty members' teaching effectiveness. With the placement of these questions, we may assume students may be summarizing the previous responses of how they would rate the course or instructor overall. Shown throughout this manuscript, there is so much more to faculty members' performance than these two questions. Yet, the SET data was available and there was an interest if the department studied exhibited similar alignment to the research. Further, there was an

interest in whether instructors' gender played a role in the responses and if students of lower or higher level courses responded to these questions differently as well.

The study consisted of 15,858 student course survey records from 327 courses in spring 2011 through fall 2015, excluding summer courses due to unique factors influencing courses and instructors such as accelerated and typically smaller courses. The 327 courses were captured with 45 different instructors teaching during the timeframe. The level of courses, in the college assessed were mostly undergraduate course responses (n = 14,453), with graduate course responses accounting for about 10% of the records (n = 1,405). Undergraduate lower level courses consisted of 100 or 200 level with undergraduate upper level defined as 300 or 400 level courses. Graduate courses were 500, 600, and 700 level courses. Assessment of the graduate courses were not included in the study. Table 8 shows the differentiation of SETs by course levels in the dataset.

Table 8. Study Dataset Detailed by Course Level Details.

Division of Course	Course Level	Number of SET records
Lower Division	100	5,071
	200	1,956
Upper Division	300	5,017
	400	2,409
	500	691
Graduate Division	600	671
	700	43
	Total	15,858

The research questions developed to assess from the SET data were:

1. What do the descriptive statistics reveal relative to how students rate the course and instructor?
2. Is there a difference between lower level classes versus higher level classes' ratings on how students rate the course?
3. Is there a difference between lower level classes versus higher level classes' ratings on how students rate their instructor?
4. Is there a difference between male versus female instructors' ratings on how students rate the course?
5. Is there a difference between male versus female instructors' ratings on how students rate their instructor?

FINDINGS

With the developed research questions, descriptive statistics and inferential statistics were used to analyze the data. Means and standard deviations (S.D.) presented the descriptive statistics. Levene's Test for Equality of Variances examined the variances between the means. A *t-test* (two-tailed) compared the difference between means for the difference research questions. Finally, differences with $p < .01$ were considered statistically significant.

The descriptive statistics explored were the collective responses to the questions 'how do you rate this course' and 'how do you rate your instructor'. The level of the course and faculty members' gender were explored for both questions of how do you rate your instructor and course to see if there were differences in students' responses.

Descriptive Statistics Rating the Course and Instructor

The first descriptive statistic assessed was the rating of the course. Students responded 78.2% of the time that the course was above average or excellent (Table 9). The mean was 4.18, S.D. = .950, with $n = 15,719$. More students responded to this question versus the how do you rate your instructor question. The mean results were lower than the scores on 'how do you rate your instructor' indicating that students rated the course lower than the instructors' ratings.

The second descriptive statistic assessed were the ratings from students for their instructor. Students responded 86.8% of the time that their instructor was above average or

excellent (Table 9). The mean was 4.49, S.D. = .813, with n = 15,651. Students rated the

instructors higher than their courses ratings.

Table 9. Responses to Survey Questions for Rating This Course and Your Instructor.

Response	How do you rate this course?			How do you rate your instructor?		
	Frequency	Percent	Cum	Frequency	Percent	Cumulative
Poor (1)	296	1.9	99.1	139	.9	98.7
Below average (2)	577	3.6	96.2	321	2.0	97.8
Average (3)	2,442	15.4	93.6	1,425	9.0	95.8
Above average (4)	5,155	32.5	78.2	3,626	22.9	86.8
Excellent (5)	7,249	45.7	45.7	10,140	63.9	63.9
Total	15,719	99.1		15,651	98.7	

The third descriptive statistic reviewed was how students rated the course and instructor

based on the level of the course. In the 15,719 records assessed for how students rated the course,

44.3% of the students were enrolled in lower level courses in the college. The mean response to

rating the course was 4.15 for lower level courses with a larger mean of 4.42 for higher-level

courses. When asked to rate their instructor, 15,651 students responded. Similarly, the mean for

the lower level courses was 4.48 with a slightly larger mean of 4.62 for the higher-level courses

(Table 10).

Table 10. Responses to Survey Questions Rating the Course and Instructor by Course Level.

Rating	Course Level	N	Mean	Std. Deviation	Std. Error Mean
How do you rate this course?	Lower	14,327	4.15	.959	.008
	Upper	1,391	4.42	.811	.022
How do you rate your instructor?	Lower	14,262	4.48	.823	.007
	Upper	1,388	4.62	.700	.019

Difference Questions on How Students Rate This Course

Assessment of whether lower level courses had different results from higher-level

courses on how students rated their courses required a Levene's Test for Equality of Variance

and *t-test* (two-tailed) analysis. The results were statistically significant. Equal variances cannot

be assumed, thus the *t* value of -11.614 was used for the question How do you rate this course

(Table 12). The *t* value was statistically significant ($p < .001$) representing there were differences between how lower level courses and higher-level courses were rated. The effect size for rating the course *d* = .031 was smaller than typical with a minimal relationship (Vaske, 2008).

Difference Questions on How Students Rate Their Instructor

A Levene's Test for Equality of Variance and *t-test* (two-tailed) analysis assessed lower level courses versus higher-level courses on how students rated their instructor. The results were statistically significant. Thus, Equal variances cannot be assumed and the *t* value of -7.196 was used for how students rated their instructor (Table 11). The *t* value was statistically significant (*p* = .000) representing there are differences between how students rated their instructors in lower level courses and higher-level course. The effect size for rating the instructor *d* = .184 was smaller than typical – representing a weaker relationship.

Table 11. Independent Samples Test Results for Rating this Course and Instructor.

Rating		F	Sig.	T
How do you rate this course?	Equal variances assumed	32.146	.000	-10.128
	Equal variances not assumed			-11.614
How do you rate this instructor?	Equal variances assumed	86.290	.000	- 6.301
	Equal variances not assumed			-7.196

Gender and Ratings Research Questions

The final descriptive statistic assessed the role gender of the faculty member plays in the two items student course survey results. Female instructors taught 24% of the courses during the time frame assessed. In Table 12, there is a larger variance of means, male = 4.15 and female = 4.27 (variance = .12), when answering how students rated their course than in the responses to how they rated their instructors, with male = 4.47 and female = 4.57 (variance = .10). For each question, female instructors were rated higher than male instructors. The difference in S.D. between male and female instructors on students rating their course (.112) versus rating their instructor (.074) were comparable.

91

Table 12. Instructor Gender Responses to Rating the Course and Instructor.

Rating	Faculty Gender	N	Mean	Std. Deviation	Std. Error Mean
How do you rate this course?	Male	12,703	4.15	.969	.009
	Female	3,016	4.27	.857	.016
How do you rate your instructor?	Male	12,629	4.47	.826	.007
	Female	3,022	4.57	.752	.014

Difference questions by instructor gender on rating their course responses.

As shown in the descriptive statistics and the research performed by MacNeil et al., there

was an interest whether instructors' gender played a role in responses to how students rated their

course. A Levene's Test for Equality of Variance and *t-test* (two-tailed) analysis showed the

results were statistically significant (Table 13). Thus, Equal variances could not be assumed and

the *t* value of -6.512 was used for the question how you rate this course. The *t* value, as

mentioned, was statistically significant ($p = .000$). The effect size $d = .131$ was smaller than

typical.

Table 13. Independent Samples Test Results for Rating this Course and Instructor based on
Instructor Gender.

Rating		F	Sig.	T
How do you rate this course?	Equal variances assumed	40.097	.000	-6.032
	Equal variances not assumed			-6.512
How do you rate this instructor?	Equal variances assumed	71.034	.000	- 6.100
	Equal variances not assumed			-6.462

Instructor gender difference questions when rating their instructor.

Knowing the results of the difference question for instructor gender for courses was

statistically significant; there was an interest whether instructors' gender played a role in student'

responses rating their instructor. A Levene's Test for Equality of Variance and *t-test* (two-tailed)

analysis assessed how students rated their instructor based on gender. These results were

statistically significant ($p = .000$). Therefore, Equal variances cannot be assumed for how you

rate your instructor and the *t* value of -6.462 was used. The effect size *d* = .127 was smaller than typical indicating a minimal relationship.

CONCLUSION

Student Evaluation of Teaching surveys currently act as an input for assessing teaching effectiveness in higher education. In addition, students provide feedback on courses and instructors via these course surveys. The instructor has the ability to influence the seven constructs identified within their courses. In the college studied, both course and instructor performance were rated above average and excellent, in 78.2% and 86.8% respectively, for the SETs reviewed. Yet, whether student learning outcomes occurred remains a larger challenge to assess. One may infer the instructor incorporated these constructs in their courses in one way or another to receive such high summative ratings.

A critical gap in higher education is identifying a way to understand the relationship of a faculty member's performance and their ability to influence student learning. The success of students' learning outcomes theoretically should be helped with the incorporation of the seven constructs in a course. Faculty members assisted in the identification of the student measured constructs and vetted the best practices in their course. Other than SETs, how might faculty members gauge student success? As Palmquist (2011) and the 2011 TILT taskforce suggested, other teaching effectiveness assessment tools are peer review, faculty reflective statements, or faculty teaching portfolios, to name a few.

Peer feedback offers a different method of formative feedback for faculty members. It may be more effective than student feedback in evaluating performance as faculty members' experiences in the classroom can offer suggestions to drive improvements for their peers. Palermo (2013) recognizes the importance of collaboration among faculty members when modifying teaching behaviors to ensure faculty members are supportive of each other. Yet the

constraints of time, faculty bias, and level of expertise for evaluation guide the literature when discussing peer review.

Faculty reflective statements represent another input of assessment performed by instructors (Palmquist, 2011) for formative assessment. These statements can capture the activities faculty members performed to ensure learning outcomes. The statements allow faculty to review what has worked and not worked, from their perspective, in their courses. As these are personal reflective statements, faculty members do not need to share the information with their peers and can develop a strategy of change for activities to improve (Seldin, 1999). The drawbacks of reflective statements are the lack of objectivity, expertise of new faculty members, and value assessed by administrators. The input of other faculty members attempts to address some of the drawbacks. Reflective statements integrated with student feedback provide additional feedback on teaching effectiveness.

A teaching portfolio is a comprehensive assessment tool for teaching effectiveness. A teaching portfolio is "a collection of carefully selected artifacts accompanied by explanatory narrative statements, which together provide evidence of teaching effectiveness" (Lattuca & Domagal-Goldman, 2007, p. 84). A teaching portfolio includes summarized SET data, peer feedback, reflective statements, and other assessment data collected by the faculty member. The teaching portfolio "features a faculty member's scholarly exploration of his or her design, development, implementation, and refinement of chosen courses" (Marincovich, 1999, p. 52). Marincovich posits the attainment of successful learning outcomes are at the forefront of faculty members' responsibilities.

SETs play a role in the assessment of faculty members' performance and influence on student learning outcomes. It is one common input of faculty members' performance. With the

various contextual factors influencing the ratings on the SETs, this study showed that lower level and upper level courses and the gender of faculty members can actually provide biased feedback. By utilizing the newly identified seven constructs to define teaching effectiveness and using varied assessment methods explored in this manuscript, teaching effectiveness can continue to play an important role in higher education.

REFERENCES

American Association of University Professors. (2010). Tenure and teaching-intensive appointments. Retrieved from http://www.aaup.org/report/tenure-and-teaching-intensive-appointments

Anderson, W.A., Banerjee, U., Drennan, C.L., Elgin, S.C.R., Epstein, I.R., Handelsman, J., … Warner, I.M. (2011). Changing the culture of science education at research universities. *Science*, *331,* 152-153.

Bain, K. (2004). *What The Best College Teachers Do.* Cambridge, MA: Harvard University Press.

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review, 31*, 709-719.

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

Boyer, E.L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate.* New York, NY: The Carnegie Foundation for the Advancement of Teaching.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review, 41*, 71-88.

Brydges, S., Chilukuri, L., Cook, G., Feeley, M., Herbst, M., Tour, E., ...Van Den Eide, L. (2012). Building a faculty learning community at a research university. *Currents in Teaching and Learning, 5*(1&2), 17-34.

Carnegie Foundation for the Advancement of Teaching. (1998). *Reinventing undergraduate education: A blueprint for America's research universities.* Princeton, NJ: Boyer Commission on Educating Undergraduates in the Research University.

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

Colorado State University – The Institute for Learning and Teaching (TILT) (2015). Evaluating teaching effectiveness task force. Retrieved from https://tilt.colostate.edu/sotl/taskForces/effectiveness/.

Corrigan, P.T. (2012). *Supporting learning: What faculty want from administrators*. In Magna Report - Administrator leadership and faculty support: Best practices for supporting faculty (4-5). Retrieved from: http://www.magnapubs.com/newsletter/academic-leader/106/supporting_learning_what_faculty_want_from_administrators-11492-1.html

Czikszentmihalyi, M. (1982). Intrinsic motivation and effective teaching: A flow analysis. *New Directions for Teaching and Learning: Motivating Professors to Teach Effectively, 10*, 15-25.

Dezure, D. (1999). Evaluating teaching through peer classroom observation. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 70-96). Bolton, MA: Anker Publishing Company, Inc.

Drennan, C.L., Elgin, S.C.R., Epstein, I.R., Handelsman, J., Hatfull, G.F., Losick, R., ... Warner, I.M. (2011). Changing the culture of science education at research universities. *Science*, *331,* 152-153.

Feden, P.D. (2012). Teaching without telling: Contemporary pedagogical theory put into practice. *Journal on Excellence in College Teaching, 23*(2), 5-23.

Frost, S.H. & Teodorescu, D. (2001). Teaching excellence: How faculty guided change at a research university. *The Review of Higher Education, 24*(4), 397-415.

Grassian, D. (2013). Constructive ambiguities: The inspiring, deflating, transformative, and limited possibilities for assessment in higher education. *Journal of Excellence in College Teaching, 24*(2), 155-173.

Holman, K. (2013). *Evaluation of student engagement assessment in Colorado State University's Warner College of Natural Resources*. Retrieved from ProQuest Digital Dissertations.

Huber, M.T. (2002). Faculty evaluation and the development of academic careers. *New Directions for Institutional Research, 114,* 73-83.

Krathwohl, D.R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 1-54.

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

Lattuca, L.R. & Domagal-Goldman, J.M. (2007). Using qualitative methods to assess teaching effectiveness. *New Directions for Institutional Research, 136,* 81-93.

Layne, L. (2012). Defining effective teaching. *Journal of Excellence in College Teaching, 23(*1), 43-68.

Lieberman, D.A. (1999). Evaluating teaching through electronic assessment. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 139-152). Bolton, MA: Anker Publishing Company, Inc.

Lumina Foundation - Degree Qualification Profile (2016). Retrieved from https://www.luminafoundation.org/files/resources/dqp.pdf

MacNeil, L., Driscoll, A., & Hunt, A.N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40,* 291-303.

Mann, K.V. (2010). Self-assessment: The complex process of determining "How we are doing" – A perspective from medical education. *Academy of Management Learning & Education, 9*(2), 305-313.

Marincovich, M. (1999). Using student feedback to improve teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 45-67). Bolton, MA: Anker Publishing Company, Inc.

McMillan, J.H. & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons, 87*(1), 40-49.

Miller, J.E. & Seldin, P. (2016). Changing practices in faculty evaluation. *American Association of University Professors*. Retrieved from: http://aaup.org/.

O'Neill, G.P. (1988). Teaching effectiveness: A review of the research. *Canadian Journal of Education, 13*(1), 162-185.

Palermo, J. (2013). Linking student evaluations to institutional goals: A change story. *Assessment & Evaluation in Higher Education, 38*(2), 211-223.

Palmquist, M. (2011, December). *University Distinguished Teaching Scholars - The Institute for Learning and Teaching Assessing Teaching Effectiveness Task Force Recommendations.* Retrieved from http://tilt.colostate.edu/sotl/taskforces/teaching/recommendations.pdf.

Pan, D., Tan, G.S.H., Ragupathi, K., Booluck, K., Roop, R., & Ip, Y.K. (2009). Profiling teaching/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Research in Higher Education, 50,* 73-100.

Porter, S.R., Rumann, C., & Pontius J. (2011). The validity of student engagement survey questions: Can we accurately measure academic challenge? *New Directions for Institutional Research, 150,* 87-98.

Rice, R.E. (2002). Beyond *Scholarship Reconsidered:* Toward an enlarged vision of the scholarly work of faculty members. *New Directions for Teaching and Learning, 90,* 7-17.

Ried, L.D. (2011). A model for curricular quality assessment and improvement. *American Journal of Pharmaceutical Education, 75*(10), 1-9.

Seldin, P. (1999). Using student feedback to improve teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 97-115). Bolton, MA: Anker Publishing Company, Inc.

Smith, D.B. & Gadbury-Amyor, C.C. (2014). Process evaluation of a teaching and learning centre at a research university. *Assessment & Evaluation in Higher Education, 39*(4), 427-442.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83(*4), 598-642.

Stark, P.B. & Freishtat, R. (2014). An evaluation of course evaluations. University of California Berkeley. Retrieved from: http://senate.csusb.edu/Reports/evaluations_paper.pdf. 1-15.

Vaske, J. (2008). *Survey Research and Analysis: Applications in Parks, Recreation and Human Dimensions.* State College, PA: Venture Publishing, Inc..

Weiman, C. (2015). A better way to evaluate undergraduate teaching. *Change.* January/February, 6-15.

Williams, K.M. (2015). *Doing Research to Improve Teaching and Learning: A Guide for College and University Faculty.* New York, NY: Routledge.

Yang, J. & Walker, Z. (2015, Aug 04). Fewer parents paying for college. *USA Today.*

EVALUATION OF TEACHING: WHAT WE LEARN FROM STUDENTS' WRITTEN COMMENTS

SUMMARY

There is a tradition in higher education to obtain feedback on the course, the instructor, and/or other selected factors from students at the end of the term. Most universities ask specific questions established by a collaborative group to collect targeted feedback. Some use Likert-scaled items formatted for the responses. Some provide a place for students to write comments. If a goal of the feedback is to drive continuous improvement in courses, the curriculum, or for the performance of the individual instructors, what information do students share in the written comments? After analyzing 679 written comment records from student course surveys, from 46 courses, over four semesters within a department at a Research I: Doctoral University, the themes identified aligned with the quantitative responses. The themes were highly positive though providing minimal constructive feedback on what is working or not working in the courses. Words such as 'well taught course with a great prof' and '… truly expressed his passion and shared it with students' were captured in the transcription process. Word clouds graphically represented the themes in the written comments with descriptive statistics used to analyze the quantitative findings to identify trends in the data.

INTRODUCTION

The act of teaching is a complex, progressive profession. Individuals' perceptions of what it takes to 'teach' 10-12 hours of college a semester is often underestimated and naïve (Cashin, 1999). When discussing the scholarship of teaching, Boyer (1990) noted the "lack of awareness of the hard work and the serious study that undergirds good teaching" (p. 23). From the development of a new course to the execution of an existing course, there are numerous steps in the process which faculty apply to maintain relevance of the course. It requires time to review

and incorporate current events and examples while concurrently adapting to a new student group each term. Articulating the activities and pedagogy, which support the processes used to develop courses, is more of an exception than the rule (Huber, 2002; Lattuca & Domagal-Goldman, 2007; Layne, 2012). Often the philosophy of a department or university will direct and dictate some of the practices. Feedback, some from student course surveys, may suggest it is necessary to update course content, delivery, and materials. These efforts may be trial and error provided in different courses and may vary in effectiveness at times. Developing the content for a course, prior to the actual delivery plays a key role in teaching effectiveness and the success of students' learning outcomes.

Student Evaluation of Teaching (SET) instruments are one of the common tools used for both formative and summative purposes of faculty performance in higher education. Some question if these are the appropriate indicators of student or faculty performance (Spooren, Brockx, & Mortelmans, 2013). Common purposes of SET's are to inform the improvement of teaching and administrative decisions used for formative assessments (Dezure, 1999; Marincovich, 1999; Seldin, 1999). Some organizations use the assessment for formative purposes, others for summative purposes (Cashin, 1999; Dezure, 1999; Ried, 2011; Zubizarreta, 1999). Summative assessments provide data for decisions about faculty members' positions, promotion, tenure, and merit increases.

The frequent availability of the SET instruments makes it difficult to substitute other indicators for assessing faculty members' performance. Yet, these course survey instruments represent one indicator of performance for a faculty member. This study searched for the type of feedback faculty members can attain from students' written comments. It posited that

constructive feedback from students' written comments on SETs may indicate teaching effectiveness for formative and summative assessment.

In higher education one could argue, assessment of performance measurement includes factors, such as grades earned by students, research publications, community involvement, and SETs. These factors exist in a system, with many departments and support staff who help students succeed. How do we gauge if faculty are successful in contributing to student learning outcomes and success in their careers? While studies are published relative to the role of student course evaluations (Braga, Paccagnella, & Pellizzari, 2014; MacNeil, Driscoll, & Hunt, 2014; Smith & Gadbury-Amyor, 2014; Stark & Freishtat, 2014; Weiman, 2015; Williams, 2015), actual uses do not appear to be consistent within the performance appraisal and/or promotion and tenure processes (Cashin, 1999; Dezure, 1999; Lieberman, 1999; McMillan & Hearn, 2008). For example, Cashin (1999) suggests the SET scores should not be averaged when reviewing faculty performance as it does not provide the range of responses and may be skewed. Yet, SETs as average scores of Likert-scaled items, from students present the day the evaluations were completed typically are used in the performance evaluations.

Lieberman (1999) further proposes Classroom Assessment Techniques (CAT) as a way to collect formative information from students to represent instructor performance. The CAT allows students to provide feedback of the effectiveness in the classroom – from the perspective of the students. However, the perception is this is 'ungraded feedback' (p. 145), which allows the integration of feedback into the next course offering to benefit students but is not as helpful to students in the current course. With so many opinions and points in the research, curiosity grew in the value students' written comments could add in improving the courses and learning outcomes versus faculty members just receiving quantitative SET scores.

There was interest in answering the question if there was better feedback from the students who wrote comments on the SETs beyond responding to the Likert scaled items to assess teaching performance. A mean, median, or standard deviation of a Likert scaled item while providing one point of feedback may not provide enough information for a faculty member to understand students' perceptions of their performance. Before delving further into the topic, we needed to articulate the relationship between teaching performance and learning outcomes to ensure student success.

Teaching Performance and Learning Outcomes

It is difficult to discuss teaching effectiveness without including a discussion about student learning outcomes. While defining how and what to evaluate, consideration must be focused on defining outcomes relative to course objectives. These objectives include student learning outcomes and student success. Pedagogical methods evolve over time, especially influenced by technology. However, what faculty members deliver during a course and students' takeaway need to be aligned.

Providing faculty members with the definition and assessment criteria for their teaching is an important step prior to any assessment. Studies have focused on assessment tools (e.g., student course evaluations, peer review, and portfolios) or topics related to teacher effectiveness or instructional quality. These studies commented on the challenges with a) assessment and defining effectiveness or quality of teaching; such as bias related to students choosing professors based on peers' recommendations (Braga et al., 2014), b) reliability of feedback (Marincovich, 1999), and c) competency of students to accurately respond to items (Beleche, Fairris, & Marks, 2012). These studies fell short of providing resources for faculty to use to assess learning outcomes in their courses. Further, in the literature reviewed, assessment and pedagogy are often considered separate topics.

Course evaluation research is a common theme and repeatedly referenced as one input toward the evaluation of teaching. Yet other resources faculty could use to help guide them toward better teaching and expected learning outcomes necessitated further research. While there is much research on SETs, students' written comments as a developmental resource for faculty members receives little attention in research. Thus this study would focus on the usefulness of written comments in improving the faculty members' courses.

The challenge in the SET ratings falls between what faculty are able to influence in a course and the perceptions of students when evaluating faculty performance. There are numerous factors already discussed which can influence students ratings on the SETs. Yet, could the written comments on the SETs provide better informative feedback versus SET quantitative scale information? If faculty successfully navigate these factors as they attempt to achieve learning outcomes in a course, is this reflected in the students' written comments and the Likert scaled items in SETs?

As Williams (2015) posited, learning outcomes are "what faculty want students to know or do as a result of the instructional experience designed" (p. 78). Departmental faculty members determine teaching decisions and learning outcomes. However, once defined the attainment of learning outcomes needs assessed through students' performance. Students' performance on course work, assignments, or exams per the course grade are sometimes used to measure faculty members' success. Interestingly, Grassian (2013) argued "exceptional teaching should not be part of a learning outcome for a course" (p. 167). He went on to highlight that the workplace should have valuable input on learning outcomes to ensure graduates are successful when entering the workforce. Lattuca and Domagal-Goldman (2007) believed any assessment of teaching should include the faculty members' ability to contribute successfully to the desired

student learning outcomes. Thus, the students' perceptions of their achievement of the learning outcomes, assessed in the SETs may represent one input to the faculty members' performance. The challenge with this perspective lies in students' ability to recognize learning and to understand the pedagogical intent of the instructor.

Given all of the decisions faculty members make before, during, and after teaching a course, learning outcomes are difficult to measure and tie to faculty performance. Mann (2010) defined two types of learning outcomes, affective and cognitive, in his study of self-assessment of all levels of employees in the medical field. Affective outcomes include reactions, motivations, and self-efficacy. This assessment includes students' satisfaction, application of new knowledge, and perceptions of their abilities. Cognitive outcomes include "understanding of task-relevant verbal information, including both factual and skill-based knowledge" (p. 306). Palmquist (2011) and the TILT taskforce argue "… what students take away from a course in terms of knowledge, skills, attitudes, and abilities – are not synonymous with teaching effectiveness. Although they are closely linked, it is possible (albeit rare) to teach a course well without necessarily achieving the learning outcomes associated with course goals" (p. 1). SET items attempt to get at both affective and cognitive learning outcomes per Mann's (2010) definition.

A convenience sample of universities and their SETs and/or teaching effectiveness questions were collected (Table 14) to benchmark teaching performance for their institution. They provided examples of wording and questions captured through their SETs instruments. Each university's SETs had items about the instructor (5) specifically, while a smaller number of them asked about the course (3) or student (3) role in learning. Given the SET items ratings

combined with the feedback from written comments, there was an interest for SETs to provide

"good and informal feedback" (Bright, 2013, para. 3).

Table 14. Institutional Benchmarking Summary of SET Examples.

Institution Name	Location	Nature of Students' Items on SET
Colorado State University	Fort Collins, CO	Course (11), Instructor (12), Student (4), Classroom and Facilities (2)
La Trobe University	Melbourne, Australia	Instructor (8), Student (6), Summary (2)
Monash University	Melbourne, Australia	Instructor (4)
Marshall University	Huntington, WV	Course (6), Instructor (16)
Towson University	Towson, MD	Course (5), Instructor (6), Student (6), Mission Driven (4)

What items are students most capable of answering on the SETs? For example, are

students capable of assessing the performance of the instructor, 'How effectively did the

instructor facilitate student learning'? It is difficult to assess, as Palmquist (2011) and the TILT

taskforce highlighted, if the students achieved the desired learning outcomes for a course. Yet, a

cumulative score may represent students' responses to the SET items. With numerous factors

influencing the ability of the faculty member to achieve the desired learning outcomes, this

question represents one aspect of the teaching versus learning debate.

Student Course Survey Organization

Student evaluations of teaching traditionally address areas relative to the course and

instructor in an attempt to measure students' perceptions of the course as noted from the five

institutions. Many of the questions are associated with topics discussed thus far, which are

essential to learning by students. Often, there may be a question about student learning. At the

studied university, course information requested at the beginning of the SET include the course

number, section number, date, and instructor's name. There are 29 Likert scaled items and two

multiple-choice questions within the 'About the Course' and 'About the Instructor' sections

addressing feedback about the course, instructor, classroom and facilities, and student (Table 15).

The third research question of this manuscript desired to assess the written comment themes

related to the Likert-scaled ratings on two questions, one rating the course and one rating the

instructor.

Table 15. Topic Areas of Student Evaluation of Teaching Items and Sample Items at Colorado State University.

Topic Area	Number of Items	Sample Item(s)
About the Course	11	How well did class sessions increase your understanding of the subject? How do you rate the course?
About the Instructor	12	How effectively did the instructor facilitate student learning? How do you rate your instructor?
About the Classroom and Facilities	2	How do you rate the overall quality of the classroom?
About the Student	4	How do you rate your level of effort in this course?
Open Ended	N/A	11 prompts for written comments

The two questions, 'How do you rate this instructor?' and 'How do you rate this course?'

used in the focus of faculty members' performance evaluations were captured in the 'about the

course' and 'about the instructor' section of the SET, respectively. The responses on these two

questions as rated provided a greater impetus to assess the written comments and leverage the

feedback from students on their courses and instructors.

Question #30 of the SET, used in this study, were the written comments transcribed for

analysis. Question #30 had a large box with eleven prompts detailed in the text as listed below:

WRITTEN COMMENTS   Please write your comments in the blank section below.

These written comments will be provided to the course instructor and may or may not be used

for the evaluation of teaching performance. You may send separate, signed comments to the department chair or head to ensure they are considered for evaluation purposes.

30. Please comment on any other items you may wish to address. Possible items for consideration include: (1) command of the subject matter, (2) enthusiasm for teaching and learning, (3) stimulation of students to do creative work, (4) effectiveness in advising, (5) effectiveness in choosing technology and providing directions for its use, (6) promotion of mutual respect in a climate free of discriminatory behavior, (7) improvement to future offerings of the course, (8) types of activities or assignments that contributed most to your experience and learning in this course, (9) types of activities or assignments that contributed least to your experience and learning in this course, (10) what you enjoyed most about the course, and (11) any general observations.

Question #30 suggests opportunities for students to write about learning and learning outcomes. However, there are many details to read before responding. After responding to 29 Likert-scaled questions, what would the written responses look like? While the instructions act as a guide to invoke student thinking on their written comments many of the prompts appear technical and intimidating: command of the subject matter, stimulation, effectiveness, and discriminatory behavior. The readability grade level for question #30, based on six different levels and indexes found at Readability (2011) is 20.9 grade level suggesting the level of understanding prompted is quite high. The range of grade levels for that readability scores were 8.9 to 31.8. Many students may not understand what each item is asking and thus struggle to provide quality written comments. Yet, after reading (or not) the instructions for question #30, 4.5% of the sample of SETs reviewed ($n = 689$) included written comments. If this is the first time a student has completed a SET in a course, they may read the instructions or run out of time

to complete; however, often students have completed SETs and may not even read the

instructions.

Student Bias in Rating Course Survey Items

An assumption is that students' voices represent important voices and opinions in

capturing faculty members' performance. The faculty-student interaction throughout a term and

students' opinions about course materials, course delivery, instructors' characteristics are some

of the questions students answer in SETs. Yet, students may not be in the best position to provide

valid answers to some of these questions. Students' knowledge about what it takes to design

and/or teach a course is limited. Given their limited knowledge of teaching, the course survey

then only asks for their opinions.

How often and to what extent is there bias in students' responses? Are they even aware of

their biases? Huber and Power's (1985) research summarized the "four primary reasons that

informants provide inaccurate or biased data" (p. 172). In this article, the informants are the

students, and their responses may not be as accurate as desired because they:

1. Are motivated to do so.
2. Have perceptual and cognitive limitations resulting in inadvertent errors.
3. Lack crucial information about the event of interest.
4. Have been questioned with an inappropriate data elicitation procedure. (Huber & Power, 1985, p. 172).

Self-report measures on course surveys pose a concern if used for faculty members'

assessment based on the potential bias of students' responses. While some fields of study can

utilize physical measures to gauge the amount of change from a program, these situations require

the participants' responses to explain *how* achievement of the change occurred. Students'

evaluation of a course is similar. With the many non-course factors students face during a course,

before a course, or even after a course, which one(s) most influences their responses? An added

difficulty to capturing self-report measures is that students may not always be aware of the full

value of a course when they are responding to the SET questions. Thus, consideration to the timing of the student course surveys distribution is important.

PURPOSE OF STUDY

By utilizing the spring 2014 through fall 2015 written comments data, the study desired to identify themes on teaching effectiveness for faculty to ensure student learning outcomes. The SETs utilized in the department of this study contained both quantitative, Likert-type items and qualitative, open-ended questions. This manuscript focuses on the following research questions:

1. How frequently did students respond to the open-ended questions providing comments in face-to-face courses?

2. What were the most common students' written comments and how could they benefit faculty?

3. What relationship was there between the written comments and the Likert-rating to the questions 'How do you rate this course?' and 'How do you rate this instructor?'

METHODOLOGY

Each SET record represented one student's course survey data from one course. There was extensive data contained within the records – course number, expected grade, faculty member identification, Likert-scaled items on evaluation of the questions, self-recorded attendance, semester, option for signature by student, and a few additional items. However, the written comments were not part of the initial dataset accessed.

The university and college codes allows "Only instructors to view scanned survey forms. Instructors can grant access to others so they can view the forms. Instructors can also revoke any access they've granted." ("college code", 2011)**.** Written comments are available to the individual faculty member for formative feedback and their personal use. To gain access to the written comment records, it was necessary to get the department head and each faculty members'

consent. A letter was prepared requesting each faculty member's release of his or her written comments from fall 2011 through fall 2015 (Appendix E). The request letter detailed their name, course details, and assured no personal identifiers would be used in any publications.

With 679 written comments captured from 46 courses during this period, analysis could provide feedback to benefit faculty members. Based on the manual dictation involved, the final dataset assessed contained the spring 2014 through fall 2015 timeframe, thus four semesters. This represented diversity in undergraduate and graduate courses, semesters, and instructors and provided a manageable number of written comments. Each written comment was transcribed manually from .pdf ® files to excel format for analysis.

Transcription Process

Given the comments were in .pdf © files, it was necessary to transcribe the written comments. Six volunteers transcribed the written comments per the transcription process criteria. An excel spreadsheet was created of the transcribed written comments and additional information from each record (Table 16 shows two transcribed examples). Added later the course, semester, and instructor's pseudonym name for each comment allowed additional demographic data, such as gender and position while protecting faculty members identity. The transcribers noted a 'yes' or 'no' if the SETs were signed by the student. Two quantitative values captured for each SET with a written comment were: 'How do you rate this course' (question #11) and 'how do you rate this instructor' (question #23). The last step transcribed the written comments from the .pdf comment box.

Table 16. Example of Variables and Written Comment Transcription Capture.

Course #	Semester	Signed by Student?	How do you rate this course? #11	How do you rate this instructor? #23	Comment
Course A	Fall 2014	Yes	4	5	... A well-educated instructor with great expertise in department. I would recommend [instructor] to all students in the department.
Course B	Spring 2015	Yes	5	5	Overall, this was a great class. The material was fascinating and I enjoyed attending class. Great class!!

Relevance and consistency of each file ensued in the final dataset captured in the transcription process. Non-use deletion of data occurred, as there were 30 instances of undecipherable handwritten comments, either a sentence or a whole comment. Transcription of these records noted 'unreadable'. There were 26 pictures drawn as a comment or part of a comment with 17 smiley faces and 5 hearts deemed representing positive feedback. There was one frown transcribed as a negative comment. An additional comment transcribed numerous times in the written comments referenced 'class' ($n = 458$) and 'course' ($n = 234$), which provided no value-add to the study. While included in the coded data file, the elimination of these references for analysis allowed the Nvivo$^©$ and word cloud software to better highlight the feedback represented in the comments.

Two additional iterations of the dataset occurred to protect the identity of each of the instructors. The first was to provide a pseudonym for each instructor for anonymity. When

reviewing the data, instructors' pseudonyms appeared frequently in the written comments. Thus,

a second iteration occurred to delete any instructor names, real or pseudonym. This helped

ensure the analysis of written comments focused on content students had written and not the

names or pseudonyms of the faculty members.

Data Checking

As a check and balance, a comparison of the transcribed written comment analytics to the

SET data records occurred to ensure the correct number of written comments for each SET

record was identified. Any comments not matching were reviewed for accuracy or deleted from

the dataset. The data compared the quantitative files to the qualitative files to ensure alignment of

course information, faculty member name, and semester taught information. The last check of

the written comments assessed spelling and clarity prior to any analysis.

Coding the Written Comments – The Process with Challenges

The initial coding of the written comments hoped to identify each of the formative or

summative feedback, favorable or unfavorable comments, and actionable or unactionable

suggestions (Table 17). It was quickly apparent though that coding these items might prove

unrealistic.

Formative and summative assessment are forms of evaluation used to assess teaching

effectiveness, yet they serve different purposes. Formative assessment focuses on improving the

performance in teaching and differs from summative assessment, often used in personnel

manners of promotion and tenure (Centra, 1987; Pan et al., 2009).

Table 17. Focus Areas for Coding Student Written Comments.

Focused Area for Coding	Purpose of Area	Examples
Formative feedback	"…used for the improvement of teaching…" (Dezure, 1999, p. 76)	Peer evaluation process
Summative feedback	"…to serve decisions or assist in making judgments" (Fitzpatrick, Sanders, & Worthen, 2011, p. 21)	Tenure portfolio
Favorable comments	Positive feedback about the course, instructor, or materials	'I learned a lot through the self-evaluation project.'
Unfavorable comments	Feedback, negative and ideally constructive of suggested changes for the course, instructor, or materials	'The instructor was confusing.'
Actionable suggestions	Feedback providing or suggesting changes the instructor can undertake	'I would like the instructor to talk louder.'
Unactionable (Other) suggestions	Feedback which may be difficult to change in order to improve the course, instructor, or materials	'I didn't like the time of the class.'

In reviewing the written comments – it was highly improbable most students utilized any of the eleven prompts and/or focused on teaching effectiveness. While they may provide formative feedback on the course or the faculty member, from their perspective, their lack of educational experience as a teacher, proved difficult for assessment. For example, a student wrote the following comment, 'I love [instructor].' What can a faculty member deduce from this comment? Should they change their presence in the course? What did the faculty member do the student liked best? Did the student learn as intended? How can other instructors be 'loved' like [instructor]? Thus, while the written comments do provide feedback, many were not formative.

When exploring the written comments to code as favorable and unfavorable, it was challenging and discouraging. It was difficult to determine if a statement was favorable,

unfavorable, or both. For example, a student wrote two sentences in their comment, '[instructor] clearly cares more about education and growth than grading assignments. We need more teachers like [instructor]!' It may seem the student appreciates the faculty member not grading assignments. The reference that 'we need more teachers like [instructor]!" seems to imply this is a favorable comment. Is it favorable? How do the students earn a grade in the course? An additional written comment blurring the lines of favorable or unfavorable comments was, 'Very impersonal and it seemed more about memorization. I liked the material!' This comment contains two sentences with different perspectives and thus separate coding – 1 favorable and 1 unfavorable. Could you categorize this comment for improvement of the course relative to student engagement or appreciation of the topic? Is there validity to the comment if a student feels something is 'unhelpful'? Another written comment provided constructive criticism on the course and ended with 'I am not a fan of [instructor name]' which again contains different perspectives and is difficult to categorize clearly whether it is favorable or unfavorable.

The last criterion proposed to use to code the written comments was whether the idea was actionable or non-actionable. If the goal of formative assessment is to improve teaching and summative assessment is to "provide information to serve decisions or assist in making judgements" (Fitzpatrick et. al, 2011, p. 21), coding actionable versus unactionable statements could provide feedback for teaching effectiveness. The goal of this coding was to determine if each of the written comments were actionable, where theoretically an instructor could make a decision to act upon the students' input. However, some of the written comments left a lot to the faculty members' imagination for the potential improvement or change. For example, 'Overall pretty good course!' or 'Weekly quizzes instead of exams are fantastic…". The first example implies, with the words 'pretty good' there were aspects of the course the student might like to

have changed. There is no suggestion on what they liked or are suggesting being changed. The second example surmises the student may not have understood the purpose of quizzes versus exams though it is clear they preferred quizzes. The ambiguity for coding whether a suggestion was actionable or unactionable would not necessarily lead to better feedback for faculty members to use to improve teaching and thus were eliminated.

There were lessons learned in establishing the criteria for coding and proceeding to try to code some sample student written comments. When determined the criteria chosen would not accurately be coded, the study evolved to look at alternative criteria for coding. Instead of trying to put the comments into a table based on the prior established criteria, word clouds presented the students' written comments for analysis to provide a visual of the written comments.

Written Comments Relationship to Learning and Teaching

Faculty in the studied department had defined seven learning and teaching areas deemed important for teaching effectiveness. The assumption was faculty members could implement the areas in a course and students could assess. The seven areas were academic challenge; student appreciation of the subject; assessment of students; community; course organization and pedagogy; student engagement; and student/instructor interaction (Table 18). Using these areas for coding the written comments proved challenging. The lack of specificity in the comments, the excessive use of adverbs, adjectives, and nouns as names, coupled with the diversity of comments within the same class prevented the coding of written comments into the seven defined areas. The goal to provide feedback to faculty on what students felt was working and not working from the comments was proving much more challenging than expected.

Table 18. Teaching Effectiveness Constructs – Applications and Examples.

Learning and Teaching Construct	Application of Construct in a Course	Examples from Students' Written Comments
Academic Challenge	Balance of not too hard and not too easy learning; rigorous enough for a sense of accomplishment	"Difficult class to understand, but 'X' does best to teach the information."
Student appreciation of the subject	Desire to learn more about the content; recommend the course to other students; choice to continue advanced classes, relevance to real world issues; enjoyment of the topic	"This class allowed me to delve deeper into subjects that I was previously exposed to"
Assessment of students	"Wide variety of methods or tools that educators use to evaluate, measure, and document the academic readiness, learning progress, skill acquisition, or educational needs of students" (Glossary of Education Reform, n.d.)	"The thought papers & comments were especially meaningful for processing what I learned throughout the course"
Community	Ability to create a collaborative learning environment with interpersonal relations (Boyer, 1990)	"… I'd really appreciate not working with the same group on both projects to get more opportunities to learn from others"
Course Organization and Pedagogy	Efforts before, during, and after a course to ensure successful learning outcomes	"Really appreciated all the thought that went into the group discussions/activities/laying out organization in advance"
Student engagement	"Degree of attention, curiosity, interest, optimism, and passion students show when they are learning or being taught, which extends to the level of motivation they have to learn and progress in their education." (Glossary of Education Reform, n.d.)	"… in how 'X' organizes the class sessions to not overwhelm you with PowerPoint slides, but focus on key concepts supported with real-life experience and relating it to current topics/debates keeps students engaged"
Student/Instructor Interaction	Amount of time and availability of instructor to meet with students	'X' was always willing to meet outside class to discuss questions and data."

FINDINGS

The SET written comments reflected the voices of the students who typically spend 10-16 weeks during a course with the faculty member. One way to capture the voice of students are the SETs. The desired outcome of this study was to explore the data in the SETs written comments from the students for the faculty members to drive effective teaching.

Initial analysis assessed the cleanliness of the data. Revisions to the dataset focused on understanding trends and themes from the student written comments.

Participants Response Rate

There were both undergraduate (524) and graduate (155) course surveys with written comments in the data. The dataset included males (6) and females (2), who were full professors (4), associate professors (2), assistant professors (1), and adjunct professors (1). Forty-six different courses were captured with some instructors teaching more than one semester, with 41% ($n = 30$) of undergraduate classes and 59% ($n = 16$) of graduate classes.

Research question #1 attempted to understand how many students responded to the open-ended questions with written comments. Student written comments counted as one response per course survey. The data set contained written comments from 77% ($n = 524$) undergraduate students and 23% ($n = 155$) graduate students. There was an increase in students enrolled in classes ranging from 628 to 862 students per semester during the timeframe assessed. This represented a 37% increase in enrollment. Enrollment information was not available for four of the classes reviewed though the written comments from these four classes are part of the study. With a slight decline in enrollment in spring 2015, the other semesters showed incremental growth with the number of students providing written comments.

The range of written comments on the SETs was from 132 to 200 during the targeted semesters (Table 19). The number of written comments did not mirror the increased course

enrollment (Figure 4). Instead, the highest number of written comments recorded in spring 2015

was 50% of the students providing comments and the lowest number of student responses in fall

2014 at 39%. The percentage of written comments closely aligned with the percentage of

enrollment numbers when assessed for undergraduate and graduate courses.

Table 19. Compilation of Semester, Course, and Instructor of Students' Written Comments.

Semester	# of Courses	# of Instructors	# of Written Comments	# of Written Comments Per Course
Spring 2014	10	7	158	15.8
Fall 2014	8	7	132	16.5
Spring 2015	12	7	200	16.6
Fall 2015	14	8	189	13.5
Total	44	29	679	15.4

Written Responses Themes

The dataset was rich with information and themes about the courses and instructors.

Research question #2 addressed the common comments from students and the benefit to faculty.

Word clouds presented these themes. The word cloud software allowed a visual representation of

the words used most frequently. The graphics displayed the themes through colorful depictions

and with the size of the word representing the frequency of use in the written comments. The

smaller the size of the font, the less frequently students used the word.

In reviewing the responses, many of the words identified were non-descriptive, making it

difficult to ascertain the intent of the students and whether actions, if any, faculty could pursue.

With faculty members reviewing these responses it could influence better learning outcomes for

students, greater student satisfaction, or higher evaluation scores. Depending on the intent for

reviewing the course survey data the emphasis of certain words would be different.
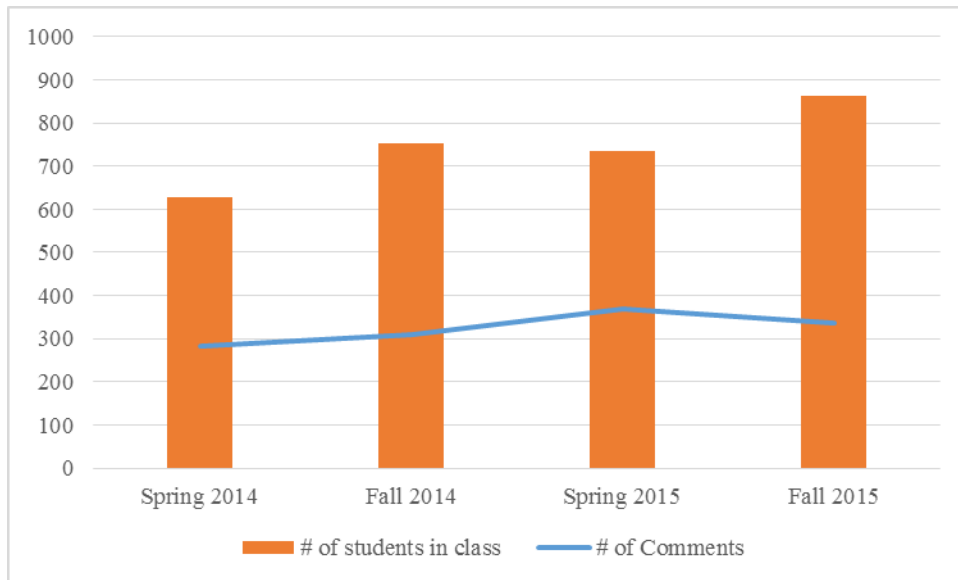
Figure 4. Written Comment Response Rate by Semester.

The written response themes represented detailed and emotional use of words by departmental, professional rank position. Often, the themes aligned to the quantitative responses. For example, '…[teacher] was very good at explaining the material and communicating to the class' received a 5 rating to the question on 'How do you rate this instructor?'

Detailed and emotional use of words by students. In reviewing the written comments, an additional observation about the emotional impact evident in the comments arose. Students' comments expressed emotions by utilizing smiley faces, capital letters for emphasis, and examples of what they liked or did not like. A sample of the emotions listed were like/liked (128), think (49), and feel (44). By utilizing phrases such as "I think" or "I feel", they clarified it was their opinion. There was a tendency for students' to be somewhat dismissive of their own opinion. For example, one student responded: "The way the groups were chosen was a little stressful and confusing, but it worked out" and another wrote "Great professor! Lots of enthusiasm! Final project has some repetitive parts that can be cut down."

When delving deeper into one of the frequently used words the word counts showed a trend. Students chose the word 'really' 90 times to add emphasis in their written comments. Thematic analysis occurred for the words immediately following 'really'. There were three different descriptors referencing 35% of the 90 descriptors (Table 20). Coding of these three descriptors assessed whether they were referring to the course, instructor or student. When referring to the course, students commented about assignments, class conditions, and what they 'liked' or 'disliked' about the course. When the descriptors referenced the instructor, comments referenced activities specific to the instructor. Comments coded as describing the student, included statements about themselves or other students in the course. So, what did the students 'really' like or dislike? And could the reference be used as indicators for continuous improvement in the course?

Table 20. Descriptor Categories and Frequencies Following the Word 'Really'.

Descriptor / Words	Expression Frequency	Word reference toward		
		Course	Instructor	Student (self or others)
Enjoyed	20	X	X	X
Appreciated/value	12	X	X	
Liked/Loved	10	X	X	
Helpful	3	X	X	
Cared	2		X	
Total	47	4	5	1

The majority of references surrounding 'really' represented positive comments. Students really 'enjoyed' the class, activities, or students. Students really 'appreciated' or 'valued' the activities, course content, and instructor. Students really 'liked' or 'loved' the learnings, the course, or the instructor. The word 'really' was used to place additional emphasis on what the students felt was 'really' good or could be improved.

Written comments – departmental professional rank positions. The words most frequently found in the dataset were great (188), learned/learning/learn (125)**,** thanks (148), really (90), and enjoyed (76). These words referenced the course, the instructor, or the students' experiences. While these words are good to see, they do not provide constructive feedback to the instructors nor do they provide clarity for any of the findings. The word cloud (Figure 5) represents the written comments for all faculty members.

Figure 5. Word Cloud Representing SETs from Comments for All Faculty Members.

An evaluation of written comments for full professors ($n = 4$) was completed (Figure 6). This word cloud reviewed various factors contributing to the potential experiences and learnings of more experienced faculty members. Full professors typically have more experience in the classroom than other ranks. The written comments could provide constructive feedback for faculty members allowing them to evaluate what is and is not working in their courses based on the students' perspective.

A comparison between the associate professor ($n = 2$) and instructor ($n = 1$) occurred. How would the experience level of an assistant professor compare to an instructor through students' written comments? There were times where some of the words appeared more frequently for assistant professors and times where the words appeared more frequently with

other levels of instructors (Table 21). For example, the word 'helpful' appeared equally in comments for courses taught. Another example, the word 'great' was used more frequently for associate professor ($n = 41$) than instructor ($n = 9$).

Figure 6. Word Cloud Representing SETs from Comments for Full Professors.

Based on written comments, there seems to be little differentiation in the written comments for courses based on faculty rank. The common words were comparable for courses taught by assistant professors ($n =1$) or associate professors ($n = 2$). Thus, while some students attempt to provide written comments for their instructors, the experience level of faculty does not seem to influence students' written comments when looking at these descriptors.

Table 21. Comparison and Descriptive Words Used in Written Comments for Emphasis.

Descriptor / word	Purpose/Explanation
Great	Referenced with Associate Professors ($n = 2$) more often
Learn	Referenced with Associate Professors ($n = 2$) more often
Helpful	Use of the word was similar for both
Assignment	Referenced with Assistant Professors ($n = 1$) more often
Good	Referenced with Associate Professors ($n = 2$) more often

Comparison to Quantitative Responses

Research Question #3 wanted to understand the relationship between the written comments and the responses to the two selected Likert-rating questions. The Likert-items are direct and offer students an opportunity to rate the course and their instructor from their perspective. The responses are based on a five-point Likert scale with the label of 5 as excellent and 1 as poor.

Table 22 shows that 83.7% of the students' SET scores rate the course as overwhelmingly positive with Above Average or Excellent ratings (Likert-item #23 on the SET). Ratings for the instructors revealed almost 90% of students rating the instructors passionately as Above Average or Excellent ratings (Likert-item #11 on the SET).

The mean score of 'how do you rate your course' was 4.26 ($n = 679$) with a standard deviation of .988 and 'how do you rate your instructor' was 4.80 ($n = 679$) with a standard deviation of .881. The students' comments were more positive when giving constructive feedback on the course. The students responded closer to Above Average (4.00) when rating the courses and closer to Excellent (5.00) when rating the instructors and providing written comments. This same tendency aligned with their written comments.

Students made suggestions of what to do more or less of and what they felt should be done differently or continued in the course. For example, "Occasionally lectures were dry. There weren't a lot of activities and [instructor] didn't move through slides very quickly. [Instructor] was definitely a valuable resource but it was hard to have back and forth between instructors. Overall great class! I enjoyed the assignments." Another example, "one of the best classes I have taken… [instructor] has great passion. I learned a lot that typical courses wouldn't be able to teach." The written comments exuded the positive attitude and experiences of the students often.

Table 22. SET Summary for Rating the Course and Instructor.

Rating Item	5 – Excellent	4 – Above Average	3 – Average	2 – Below Average	1 – Poor	N/A or Blank
The Course (q23)?						
# of responses	777	311	131	47	27	6
% of responses	59.8	23.9	10.1	3.6	2.1	.5
Total % of responses	83.7%		10.1%	5.7%		
Your Instructor (q11)?						
# of responses	983	183	82	31	16	4
% of responses	75.7	14.1	6.3	2.4	1.2	.3
Total % of responses	89.8%		6.3%	3.6%		

Further, 37.0% of the written responses were for one instructor who taught 15% of the courses (Table 23). Based on the comments, the instructors teaching style seems to have been refined over time. Students who provided written responses about instructor had strong emotional and passionate comments. For example, "Everyone who has the opportunity to take this class should. Well taught course with a great prof." and "Instructor was great at reaching out to students, and making an interesting environment. Truly expressed his passion and shared it with the students."

Overall, the comparison of the written comments with the quantitative responses rating the course and the instructor aligned. There was positive feedback for the faculty members displayed in the quantitative and qualitative responses. When the students rated the course or instructor below average (3.00) or poor (2.00), the written comments' themes similarly reflected disappointment in the course. For example, a student rated the course a 2 (of 5). When providing a written comment, the student stated "this course is a hard course to concentrate in and enjoy thoroughly. I honestly think that you need to provide better guidelines for the final project and facilitate groups better." Another student rated a different instructor a 2 (of 5) and stated "…

would be a good undergrad teacher but we felt as though [instructor] style of teaching was not effective for a grad program." These types of comments were specific and could allow individual instructors to re-evaluate the course and its delivery for improvement. However, these are singular comments from only two students in each course. Taken out of context, it would be challenging to determine the relevance of the comments for the other students in the course.

Table 23. Ratings for Course and Instructor and Frequency of Written Comments by Faculty Members.

Pseudonym Initials	Number of Courses	Average Course Rating (q23)	Average Instructor Rating (q11)	# of Written Comments	Written Comments $\bar{x}$	% of Total Courses (Dept. Studied)	% of Total Written Comments
L.B.	7	4.83	4.97	275	39.3	7.8	21.4
T.C.	6	4.11	4.35	103	17.2	6.7	8.0
L.S.	6	3.56	3.96	93	15.5	6.7	7.2
J.M.	7	4.51	4.76	68	9.7	7.8	5.2
D.M.	7	4.31	4.56	54	7.7	7.8	4.2
S.D.	5	4.18	4.70	40	8.0	5.6	3.1
J.B.	4	3.89	4.26	38	9.5	4.4	2.9
M.D.	4	4.59	4.84	37	9.3	4.4	2.8
Overall	46	4.26	4.80	708			

DISCUSSION

The students' written comments and ratings taken together on the two selected items' help provide one form of feedback. Some of the comments provided constructive feedback allowing a faculty member to see what the students liked or disliked about the course or their teaching. This study confirmed the written comments aligned with the quantitative responses. When comparing the written comments and the quantitative ratings, there were occasions where a 'great instructor' was noted and rated a 4.83 (excellent), a 4.31 (above average), and a 3.31 (average). It is difficult to gauge the differences of ratings of a 4.83 and a 4.31. Yet, the written comments can help to enlighten the instructor by adding clarity to the Likert-responses.

The written comments reflect students' perceptions. Just as there are many different types of learning environments: large lectures, labs, recitations, and seminars, and students' learning styles, it can be difficult for faculty members to apply the opinions from the courses in their designated environments. An additional challenge for faculty members to be able to act upon written comments is students are not versed in the best practices for delivery of course materials and learning. For example, if a student excelled in high school in a small class format, they may want their college classes to emulate a similar environment. A faculty member who adapts a course using a flipped classroom for example, may run into resistance as this may be different from students' experiences in the past. These factors provide insights for faculty members to the challenges of using the written comments to understand what is going well and what students believe could be improved. Asking students for improvements during the course may provide opportunities for improvements during the semester versus the written comments provided after the course is completed.

The mean scores for the courses (4.26) and instructors (4.80) items based on 679 records over four semesters in this department provided insight to the satisfaction of the students. However, the positively skewed scores delivered the message that *everyone* who teaches in the department included in the sample, across the 46 classes received an above average (4.0) rating. Table 23 shows this is not the case. L.B.' high scores and disproportional number of written comments ($n = 21.4\%$) are skewing the data. And with the assessment of SET scores, administrators must recognize skewed data could be influencing departmental scores. There are additional factors, which could influence the scores: the enthusiasm of the instructor, the students' excitement for their major and the department, the professional development opportunities faculty pursue, and the grading structures (i.e., allowing students to keep their GPA

up, grade inflation). If the SETs are the only input the department utilizes to evaluate faculty members' teaching, how are administrators determining if faculty members are effective teachers or achieving student learning outcomes?

In reviewing the 679 students' comments (29 had been removed due to being unreadable), it seemed the majority of students did not address the eleven prompts provided in the long paragraph. This made it difficult to categorize the written comments as actionable or non-actionable. Students' use of adverbs, adjectives, and nouns may not lend comments to be actionable. For example, '[instructor] is a very sweet and understanding professor making [pronoun] a great person to learn from…' is a difficult comment to code. 'Sweet and understanding' are vague relative to student learning outcomes. They do not provide context as to the achievement of learning outcomes nor how faculty members' teaching style possibly helped support the learning objectives. While there is an opportunity to inform students on how to give specific and concise feedback – suggesting they utilize verbs and examples, training on how to add value to SET feedback could lead to actionable written comments for faculty members in sustaining or improving courses.

RECOMMENDATIONS

The student course survey ratings and written comments should play a role in the evaluation of faculty, as they represent the students' voice and one factor in the assessment process. There are other ways to capture students' voices to inform teaching effectiveness, such as letters, interviews, and performance in future classes. It is important for departments, colleges, or universities to clearly articulate the purpose of teaching performance evaluation and ensure the recognition of students' voices for student success. Are students' voices read as formative feedback and an attempt to improve the performance of faculty members or are they used as a summative measure, as a factor in performance decisions? Utilizing only the student course

surveys may not be an adequate picture of the faculty members' teaching or instructional performance.

The culture of the department and each course represents context for the process of evaluation. Quality of teaching may be a priority in the department. Professional development initiatives may be encouraged. Teaching does not come naturally for everyone. In the absence of mentoring or coaching, newer teachers may replicate what they have experienced. Sometimes unknowingly, a new teacher may utilize techniques, which have been ineffective for successful learning outcomes.

Thus, substantive evaluation is critical to improving student learning, which needs to consist of varied options, especially given the context and variability of topics within departments within university settings. Faculty members should be required to submit varied, evidence-based examples of their teaching effectiveness. There are several generally accepted, best practices of deliverables to assess performance. A few examples of teaching evidence are peer reviews, reflective statements on teaching and student learning, professional development plans/trainings, mid-semester and/or periodic reviews, teaching portfolios, statements of teaching philosophy, and/or course development and design including example syllabi or evaluation of students' work (TILT, 2011). There is not a one size fits all model when it comes to evaluating teaching effectiveness. Faculty members acknowledge written comments incorporated into many of the teaching evidence examples above may promote successful student learning outcomes.

Finally, based on the written comments reviewed, there seems to be an opportunity to guide students on the purpose and value of their feedback. Many students are providing feedback aligned with the Likert-rating questions with guidance and the use of examples to demonstrate actionable feedback. If students clearly articulate specifics that worked or did not work for their

learning style in a class and then offered suggestions on what might be done to leverage or improve the course, the usefulness of the written comments would be enhanced. While it would be an individual perspective, it could be more constructive than the written comments found in this study. It would be good to proactively share with students the reasons for the various activities in the course and ask for feedback throughout the term. The student course survey should not be the first time the instructor or students respond to what is or is not working well in the course. This does not give the student, nor the instructor, an opportunity to address the situation during the class.

There is much history of SETs in higher education. The SETs represent one form of feedback, the students', responding to the course, the curriculum, the faculty member, and the experience of the student. When courses and instructors' ratings are Above Average (4.0) to Excellent (5.0), there are missed chances to parse out opportunities for improvement. Educating students on providing constructive feedback and selecting additional methods of feedback could provide a more thorough summary of the faculty members' performance. The study successfully supported answers to the research questions of how often students responded with written comments, the most common students' written comments, and the alignment from written comments and the Likert-rating.

LIMITATIONS

The study occurred in one department, using only face-to-face course data. Future research could add to the study's findings of written comments and Likert-rating questions relationship at other universities or within other departments. Expanding the research beyond one department and one college, increasing the sample size of courses, quantity of written comments, and faculty members, and assessing the quality of written comments in online courses all could shed more light on teaching effectiveness assessment.

There were significant constraints to coding the written comments for analysis. As shared, the initial categories of interest for coding were formative or summative feedback, favorable and unfavorable, actionable and unactionable suggestions. There is value in better defining criteria that would allow for a more methodical coding of the students' written comments such that faculty members understand what is or is not going well in the class.

The findings on the SET relative to the Likert-ratings on the course and instructor aligned. While the students in this department are grateful for the expertise and personalities of the instructors, it was often difficult to review teaching effectiveness when students' ratings considered so many faculty members Above Average (4.0) to Excellent (5.0). There are opportunities to educate students on the purpose of the evaluations; however, caution would be needed to ensure the principle of 'do no harm' occurred with faculty members.

REFERENCES

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review*, 31, 709-719.

Boyer, E.L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate.* New York, NY: The Carnegie Foundation for the Advancement of Teaching.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review, 41*, 71-88.

Bright, L. (2013). How do you make a 'whole person'? Retrieved from: www.beforeitsnews.com/military/2013/12/how-do-you-make-a-whole-person-2457204.html

Buehner, C.W. (1971). They may forget what you said, but they will never forget how you made them feel. Retrieved from: www.quoteinvestigator.com/2014/04/06/they-feel/

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

Centra, J.A. (1987). Formative and summative evaluation: Parody or paradox? *New Directions for Teaching and Learning*: *Techniques for Evaluating and Improving Instruction*, *31,* 47-55.

Colorado State University Academic Faculty and Administrative Professional Manual (2013). Retrieved from http://facultycouncil.colostate.edu/faculty-manual-section-e/#E.12.1

Colorado State University [university] college code (2013). Retrieved from provost.colostate.edu/codes/

Colorado State University [anonymity] college code (n.d.). Retrieved from provost.colostate.edu/codes/

Dezure, D. (1999). Evaluating teaching through peer classroom observation. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 70-96). Bolton, MA: Anker Publishing Company, Inc.

Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines.* Upper Saddle River, NJ: Pearson Education, Inc.

Glossary of Education Reform (n.d.). In *The Great Schools Partnership, the Glossary of Education Reform*. Retrieved from http://edglossary.org/

Grassian, D. (2013). Constructive ambiguities: The inspiring, deflating, transformative, and limited possibilities for assessment in higher education. *Journal of Excellence in College Teaching, 24*(2), 155-173.

Huber, M.T. (2002). Faculty evaluation and the development of academic careers. *New Directions for Institutional Research, 114,* 73-83.

Huber, M.T. & Power, D.J. (1985). Research notes and communications: Retrospective reports of strategic-level managers – guidelines for increasing their accuracy. *Strategic Management Journal, 6*, 171-180.

Lattuca, L.R. & Domagal-Goldman, J.M. (2007). Using qualitative methods to assess teaching effectiveness. *New Directions for Institutional Research, 136,* 81-93.

Layne, L. (2012). Defining effective teaching. *Journal of Excellence in College Teaching, 23(*1), 43-68.

Lieberman, D.A. (1999). Evaluating teaching through electronic assessment. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 139-152). Bolton, MA: Anker Publishing Company, Inc.

MacNeil, L., Driscoll, A., & Hunt, A.N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*, 296-303. Published online: 05 December 2014.

Mann, K.V. (2010). Self-assessment: The complex process of determining "How we are doing" – A perspective from medical education. *Academy of Management Learning & Education, 9*(2), 305-313.

Marincovich, M. (1999). Using student feedback to improve teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 45-67). Bolton, MA: Anker Publishing Company, Inc.

McMillan, J.H. & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons, 87*(1), 40-49.

Palmquist, M. (2011, December). *University Distinguished Teaching Scholars - The Institute for Learning and Teaching Assessing Teaching Effectiveness Task Force Recommendations.* Retrieved from http://tilt.colostate.edu/sotl/taskforces/teaching/recommendations.pdf.

Pan, D., Tan, G.S.H., Ragupathi, K., Booluck, K., Roop, R., & Ip, Y.K. (2009). Profiling teaching/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Research in Higher Education, 50,* 73-100.

Readability-Score. (2011). Measure the Readability of text! - improve your writing and your website marketing with Readability-Score.Com. Retrieved from https://readability-score.com/text.

Ried, L.D. (2011). A model for curricular quality assessment and improvement. *American Journal of Pharmaceutical Education, 75*(10), 1-9.

Seldin, P. (1999). Using student feedback to improve teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 97-115). Bolton, MA: Anker Publishing Company, Inc.

Smith, D.B. & Gadbury-Amyor, C.C. (2014). Process evaluation of a teaching and learning centre at a research university. *Assessment & Evaluation in Higher Education, 39*(4), 427-442.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83(*4), 598-642.

Stark, P.B. & Freishtat, R. (2014). An evaluation of course evaluations. University of California Berkeley. Retrieved from: http://senate.csusb.edu/Reports/evaluations_paper.pdf. 1-15.

Weiman, C. (2015). A better way to evaluate undergraduate teaching. *Change.* January/February, 6-15.

Williams, K.M. (2015). *Doing Research to Improve Teaching and Learning: A Guide for College and University Faculty.* New York, NY: Routledge.

Zubizarreta, J. (1999). Evaluating teaching through portfolios. In P. Seldin (Ed.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 162-182). Bolton, MA: Anker Publishing Company, Inc.

CHAPTER 5: FINDINGS AND CONCLUSIONS

INTRODUCTION

Higher education is tasked to 'create new knowledge' (Carnegie Foundation for the

Advancement of Teaching, 1998, p. 2). Student success in higher education requires faculty

members to balance numerous activities. There is the balance of students' needs, parents'

desires, administrators' dictates, and faculty members' interest in seeing students realize the

learning outcomes for a course. Defining and assessing teaching effectiveness continues to be a

difficult task with a need to articulate what is working and what could or should be improved in a

class. This dissertation presents three manuscripts aimed at ways to define and assess teaching

effectiveness. First, there was the manuscript with the exploration of a case study of a

departmental process of defining and assessing teaching effectiveness. The next manuscript

focused on the joint development and process of creating seven constructs deemed important to

measure teaching effectiveness. The development and process was coupled with the analysis of

quantitative results from student course surveys. Finally, themes identified from students' written

comments and quantitative responses to student course surveys were compared to determine if

there was richer feedback provided by students' written comments. Each manuscript provided

suggestions to assist in assessing teaching effectiveness and ensuring student learning outcomes.

Boyer acknowledged "Conditions in higher education have changed significantly in

recent years" (Boyer, 1990, p. 1) and change continues today. Faculty members experience

changes and continue to evolve their teaching practices frequently to support successful student

learning outcomes. Change can be purposeful. With the influence of technology, the changing

delivery modes of education, and the evolution of the needs of students, higher education

continues to be a focus area to meet the needs of all of the stakeholders (Lumina Foundation,

2016). A challenging feat to achieve due to the pace of change and an area getting much focus in the media and educational literature ensuring the discussions drive success for all those involved. Williams (2015) suggests "…faculty be part of a community of change rather than an isolated change position" (p. 2). The changes in teaching will continue to evolve and making faculty part of the solution will help achieve learning outcomes for students.

A factor recognized by many universities to assess teaching effectiveness are SETs. The university studied utilized questions in the context of 'about the course" and 'about the instructor' to assess teaching effectiveness. These two questions summarized the effectiveness of faculty members in two questions: 'How do you rate this course?' and 'How do you rate this instructor?' Yet, many universities are incorporating into their university code that 'the student course surveys must be used ONLY in conjunction with other sources of evidence' (Academic Faculty and Administrative Professional Manual of Colorado State University, 2018, p. 196).

Faculty members prepare, deliver, and assess a course and yet this work is often unknown and unfamiliar to students. Students are not aware of the decisions faculty members review for each course. Faculty assess what worked in a prior course to keep and what may not have been as effective and needs to be changed the next time they teach the course. This requires the consideration of other variables such as class size, setting, and time of day taught. Changes, improvements and/or modifications may come from SET feedback, faculty reflective statements, and/or the experiences of expected versus actual learning outcomes. So, how can faculty assess the steps required to teach a course? There may be faculty member induced changes and the influence of other stakeholders in implementing course design and execution. The preparation and delivery of teaching a course requires faculty members to balance many factors – some from student input and others, which students may not be aware of or understand.

The process explored in the case study identified seven constructs: academic challenge, student appreciation of the topic, assessment of students, community, course organization and pedagogy, student engagement, and student/instructor interaction believed instrumental in student success and learning outcomes. Each of these constructs identified areas students could best assess based on their knowledge and experience in a course. Building on each of the objectives of the teaching effectiveness project, these constructs were determined to be pivotal in defining and assessing teaching effectiveness.

REVISIT MANUSCRIPT FINDINGS – A CASE STUDY

A large, collaborative project such as 'Defining and Assessing Teaching Excellence in Higher Education' allowed much reflection on teaching, expectations, and learning outcomes among faculty and the advisory committee. It was a process that I have since realized mirrors the experience of faculty members when they are teaching. There are times when things are going smooth and to the expectations of the students affected and there are times when progress seems to be lacking. Gaining consensus has always been a vital part of collaborative processes. Similarly, in teaching, what happens when students are learning at different paces or come in to the class with different preparation – how should faculty members ensure successful learning outcomes?

The teaching effectiveness project attempted to define and assess teaching effectiveness for new higher teaching load positions. It involved conducting interviews, developing two quantitative instruments, and gaining consensus among departmental faculty and administration. The project provided numerous opportunities for learning and worked to identify the challenges in defining and assessing teaching effectiveness.

The first area of focus expanded to define and assess teaching effectiveness for all faculty in the department. The consensus was there was no reason to limit teaching effectiveness to the

137

new positions – all faculty members should strive for successful learning outcomes via effective teaching. This shifted the focus of the project away from establishing expectations for new positions to include expectations for all faculty members and creating a new vision of teaching effectiveness in the department.

The second area of focus was the creation of the various qualitative and quantitative instruments. The development of the interview instrument occurred during the early part of the project. There had been minimal research collected and reviewed. It was a compilation of 'curiosity questions' the advisory committee generated to define effective teaching. The selected questions evolved from an informal brainstorming session and were polished in future advisory committee meetings. The desire to collect the perspectives of faculty members and administrators while moving the project forward remained a priority. The pilot interview helped identify gaps in the interview instrument, which dictated revisions and the ability to capture valuable data through the remaining interviews.

The greatest challenges for the project occurred during the creation of the quantitative instruments to identify constructs faculty members believed contributed to successful learning outcomes for students. The advisory committee meetings presented options to faculty members and gained momentum as the project progressed. However, constructive feedback and questions were lacking. When beginning the design process, multiple times enthusiastic suggestions did not develop into any type of informative feedback. For example, the wording for one of the constructs evolved from "Course content" to "Organization of course content and pedagogy" to "Course organization and pedagogy". The words and intent were quite similar ensuring the incorporation of the voice of each faculty member and helped drive better buy-in for the final instrument. Yet the suggested changes to the wording of the questions was semantic. The initial

structure of the instruments were weak. Thus, time was lost trying to incorporate everyone's

suggestions into the content of the instrument. Eventually, identified selection criteria allowed

the project to move forward with a stronger instrument and consensus from the faculty members.

FINDINGS FROM RESEARCH QUESTIONS

The second and third manuscript had research questions targeted to utilize data from the

university-issued SET records. The summarized findings of each of the research questions from

each manuscript are below.

Students' Evaluation of Teaching and Beyond: One Experience – How Will You Use Them?

When asked 'how do you rate this course', 78.2% of the students responded with above

average or excellent responses ($n = 15,858$). More students rated their instructors average or

excellent than they rated the courses. It seems the dynamics and work of the instructors exceeded

the expectations students had for the courses they completed. We can infer other factors

influenced the ratings on the course – such as time of the course, interest in the topic of the

course, and a required versus elective course.

The second descriptive question addressed how students rated their instructors. 85% of

the responses rated their instructors as above average or excellent ($n = 15,848$) when asked 'how

do you rate your instructor'. The high ratings infers students were content with the performance

of their instructors.

The third descriptive focus utilized the level of the course to answer the questions 'how

do you rate this course' and 'how do you rate your instructor.' With 44.3% of the students in

lower level courses, the mean response was 4.15 and 4.48 for the questions on the course and

instructor. The high ratings infers students felt their course and instructors for all levels of

courses were above average or excellent.

The last descriptive statistic focused on the instructors' gender and the students' ratings on the questions about the course and the instructor. The gender of the instructors did not seem to influence student responses on either questions. The question 'how do you rate your instructor' resulted in a mean score of 4.15 for male and 4.27 for female instructors. The lower means for male instructors (n = 12,703) surprised the team as there are both more male instructors in the department and there were over four times more SETs assessed for male instructors than female instructors (n = 12,629).

The second research question reflected students' ratings between lower level and higher level classes' on the question rating the course. While statistically significant, the *t-test* value was -11.6141 (p < .000).

The third research question reflected students' ratings difference between lower level and higher level classes' for the question rating the instructor. While statistically significant, the *t-test* value was -7.196 (p = .000).

The fourth research question reflected students rating the difference between male versus female instructors for rating the course. Though the mean difference was small, the results were statistically significant, with the *t-test* value of -6.512 (p = .000).

The fifth research question reflected students' ratings of male versus female instructors on rating their instructor. The results were statistically significant, the *t-test* value was -6.412 (p = .000).

While the results were statistically significant, it was challenging to determine how instructors could use the findings from these questions. There were more questions than answers. Students appear to intertwine the experience of the course and the instructor. The instructor may influence students' experience in the course. In addition, factors outside of instructors' control

influences students' interest in the content of the course. The data from the descriptive questions and research questions serve as a baseline for instructors by course. When instructors agree with opportunities in the course, they will make them. Where they deem they are successfully fulfilling the objectives of the course, they should continue to teach the course in the mode they have been doing.

Evaluations of Teaching – What We Learn from Students' Written Comments

The first research question addressed how frequently students responded to the open-ended questions and provided comments from face-to-face courses. The analysis of written comments from spring 2014 through fall 2015 (four semesters) occurred. Six hundred seventy-nine written comments with a mean of 15.4 written comments per course were captured from 46 courses and 8 instructors.

The second research question addressed what the most common students' written comments were and how they may benefit faculty within the department. The elimination of words in the written comments were based on their lack of relevance. Names and non-descript words, such as really, course, and class were deleted from the data set to highlight pertinent words and feedback. The word 'really' appeared 90 times as students put emphasis on their comments. Analysis ensued to assess the words immediately following 'really'. There were 14 different descriptors accounting for 80% of the 230 descriptors identified. Further clarifying instructions to students explaining the purpose of their written comments may yield feedback with actionable words providing greater direction to the instructors.

The third research question addressed the relationship between the students' written comments and the Likert-ratings to the questions 'How do you rate this course?' and 'How do you rate this instructor?' While the departmental data reviewed does show a positive correlation of the tone of the written comments and the quantitative Likert-rating, it is difficult to know what

to do with that data. The mean scores for the courses (4.26) and instructors (4.80) items based on 679 records were extremely high and skewed from a normal curve. The data shows that the courses and faculty members teaching in the department studied were rated above average (4.00 courses) and closer to excellent (5.00 instructors) for the timeframe over four semesters studied. Further, the higher mean scores for 'how do you you're your instructor' reflects students who provided written comments, within this department, rated their instructors high.

SIGNIFICANCE AND CONTRIBUTION OF THIS STUDY

This study proved to be an inquisitive and significant learning experience utilizing student course survey data and delving into the research on teaching effectiveness in higher education. It provided an insider view on the workings of the teaching system validated through the iterative steps with faculty and administrators throughout the project. With a steep learning curve and endless question and answer sessions, my knowledge increased significantly about teaching effectiveness. From the varied use of SETs to the additional measures available to assess teaching excellence, I am confident I have just touched the surface of the topic. This dissertation is adding to the body of knowledge on teaching effectiveness.

The identification of seven constructs defining effective teaching represented an opportunity to focus on areas to increase student learning outcomes. These constructs provided items for faculty members to use in their courses. With feedback from the departmental faculty, the seven constructs developed were academic challenge, student appreciation of the subject, assessment of students, community, course organization and pedagogy, student engagement, and student/instructor interaction. The literature supported these constructs as helping students achieve successful learning outcomes. Focus on these areas in a course provides direction for faculty members and allows students to provide feedback. By focusing on these areas of teaching

effectiveness in the design, implementation, and assessment of courses, faculty members can strive for successful student learning outcomes.

Tied in with the student feedback on teaching effectiveness, this study strives to bring attention to the need for clearly articulated criteria for faculty members' performance. Teaching is part of a larger system with numerous factors influencing performance. Thus, defining and assessing teaching effectiveness becomes an integral part of faculty members and administrators' roles. While SETs represent one source of data and evidence, there are other tools and methods reviewed in the research. For example, peer reviews, faculty reflective statements, and/or teaching portfolios to name a few represent viable methods for assessing teaching effectiveness. There are numerous resources in the higher education system, which can aid in providing feedback for enhancing student learning outcomes.

REFLECTIONS OF MYSELF AS A RESEARCHER

The experience of completing a doctoral degree provides many opportunities to explore securities and insecurities as a student. Lifelong learning through my professional work, non-profit work, and other opportunities I experienced have influenced my desire to pursue a doctoral degree. I had personally led various teams to achieve successful outcomes, solved problems, executed improvement programs, and participated in numerous change initiatives. Thus, I was ready to leverage the theories and research from the program to organizations to further my impact.

At the beginning of the program, I was at the top of my professional career. A mature, successful female manager in the high tech, semiconductor industry, managing an international supply chain team, things were going well. I was balancing my professional career, non-profit volunteering, and believed I would successfully complete another graduate degree. After a merger and acquisition followed by the reallocation of resources in my professional world,

starting over in academics proved a humbling experience. My knowledge of academia was limited to my experiences as a student. The intricacies, hierarchy, and workings of academia forced me to rethink my perceptions and beliefs of higher education.

I was naïve to faculty members' experiences. From the criteria for promotion and tenure to the career paths of faculty members, there was much to learn. Addressing my instructors as 'professor' throughout my higher education experiences showed my lack of knowledge of the academic hierarchy. I strived to remain objective as I learned more and more about faculty members' roles and responsibilities. Initial journal articles I read began to shed light on the situation; however, I did not understand the 'politics' of higher education. Politics are everywhere in every work environment – why would I believe it did not exist in higher education? There was so much to learn; often times making me feel inept. I was shocked to learn through the interviews and research, how much time faculty members devoted to teaching and their courses. During the process, I shifted from recognizing what was 'broken' and needed to be 'fixed' to acknowledging the experiences in the teaching profession and the multiple pressures on faculty members. While continuous improvement had been common in my professional life, the connotation of the word, relative to teaching implied something needed improvement without thought to recognize what was going well. I shifted my work to look at resources needed to provide feedback of any kind, what was working, to improvement opportunities, instead of focusing on faculty members' performance.

IMPLICATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This teaching effectiveness study was performed in a college at a Research I: Doctoral University as a case study. There are various opportunities to perform continuing research on teaching effectiveness because there are other departments and areas focused on this relevant topic. Other colleges within the university could do their own investigation of teaching

effectiveness; leveraging the learnings from this project. Pedagogical techniques and/or learning outcome comparisons could provide additional opportunities for varied methods in courses to ensure optimal student learning outcomes. Additionally, more research at other colleges and universities could provide additional opportunities for student learning and reinforce the findings of this study.

The seven constructs identified as areas students could best assess were academic challenge, student appreciation of the topic, assessment of students, community, course organization and pedagogy, student engagement, and student/instructor interaction. Deemed important for successful learning outcomes these are areas deemed measurable by students. Testing and implementing the SET created during the TEP project would help validate if the TEP was on the right track. The goal of the instrument was to allow students to better assess their instructors based on these constructs thus positively impacting learning outcomes. Coupled with better-articulated instructions for written comments, the feedback on the constructs plus the written comments could provide faculty members with what is and is not working well in their courses. The newly developed student instrument, utilizing the seven constructs would provide both formative and summative feedback to faculty members.

Students have differing perspectives of the classes they are taking based on the subject, instructor, and perceived level of difficulty. The differences between behavioral sciences and science, technology, engineering, and math (STEM) courses may provide different perspectives on what makes a faculty member effective. Assessing the SETs based on discipline specific questions would provide more insights to what works in a class. It could highlight different delivery methods, which work best in a discipline based on the type of course taught. And

faculty members who successfully achieve student learning outcomes in one disciplines' courses could provide suggestions for other disciplines' courses.

Another area of research is to define and assess teaching effectiveness in online courses. There is an abundance of research done for online courses, which differs from the face-to-face research assessed. Online course responsibilities can be similar to face-to-face courses, such as engagement, technology, and learning outcomes. However, the fact students are not in the classroom with faculty members challenges the validation of students' experiences. Based on the growth of online degrees and courses, research continues to determine the extent to which students are achieving the skills needed to be successful in their careers. Additional online course SETs research could focus on which skills are important for the students' successful learning outcomes. Researchers could identify skills or techniques effective in both face-to-face courses and online courses to ensure student success.

SUMMARY AND CONCLUSION

The purpose of SETs, to give input to measuring teaching effectiveness will be a topic in higher education for years to come. While an easily implemented tool ingrained in the culture of many institutions, SETs allow students to provide feedback on faculty members and courses. Yet, they fall short of actually providing actionable feedback. Summarizing the quantitative ratings or averaging the responses from a course may not actually reflect the teaching effectiveness. Utilizing the average scores for two questions, 'How do you rate this course?' and 'How do you rate this instructor?' may misrepresent the actual teaching effectiveness of faculty members. There are numerous influences which impact teaching effectiveness such as, student bias, Likert-ratings, or factors influencing how a course is taught, to name a few. Some of these factors, beyond an instructors' control, influence the teaching of a course -- time of day, room

selection, and technology; SET quantitative results should take into consideration the impact of these factors.

Written comments have the ability to provide feedback on what is working well or not in a course. The instructions are critical for students to be able to provide the best feedback. Unfortunately, in the college studied, listing 11 items for written feedback does not necessarily yield constructive feedback. When the Likert-ratings align with the tone of the written comments, at least there is a level of consistency identified. However, actions faculty members can undertake to improve or continue in their courses are limited from the written comments. There is a lot of opportunity though to expand this aspect of the SET. Providing better prompts for students and providing 'good' and 'less good' examples could improve the utility of the written comments.

As stated at the beginning of the dissertation, the act of teaching is a complex, progressive task. The various stakeholders and complexity of the higher education system instill additional challenges for faculty members. Yet, students continue to graduate from college and make their marks on the world. They are learning! There are some colleges doing better at ensuring student learning outcomes than other colleges do, courses better than other courses, and faculty members better than other faculty. If we can leverage the practices of effective teachers to ensure successful learning outcomes, how much better could students do upon graduation and in college? The sky is the limit.

REFERENCES

Academic Faculty and Administrative Professional Manual of Colorado State University (2018-2019). Retrieved from http://facultycouncil.colostate.edu/faculty-manual-3.

Boyer, E.L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate.* New York, NY: The Carnegie Foundation for the Advancement of Teaching.

Carnegie Foundation for the Advancement of Teaching. (1998). *Reinventing undergraduate education: A blueprint for America's research universities.* Princeton, NJ: Boyer Commission on Educating Undergraduates in the research university.

Lumina Foundation – Degree Qualification Profile (2016). Retrieved from https://www.luminafoundation.org/files/resources/dqp.pdf.

Williams, K.M. (2015). *Doing Research to Improve Teaching and Learning: A Guide for College and University Faculty.* New York, NY: Routledge.

COLORADO STATE UNIVERSITY PERFORMANCE EXPECTATIONS FOR TENURE,
PROMOTION, AND MERIT SALARY INCREASES – TEACHING AND ADVISING
As retrieved from http://facultycouncil.colostate.edu/faculty-manual-section-e/#E.12.1.

E.12.1 Teaching and Advising *(last revised August 2, 2013)*

As part of its mission, the University is dedicated to undergraduate, graduate,
professional, and continuing education locally, nationally, and internationally. Toward
that end teachers engage learners, transfer knowledge, develop skills, create opportunities
for learning, advise, and facilitate student academic and professional development.

Teaching includes, but is not limited to, classroom and/or laboratory instruction;
individual tutoring; supervision and instruction of student researchers; clinical teaching;
field work supervision and training; preparation and supervision of teaching assistants;
service learning; outreach/engagement; and other activities that organize and disseminate
knowledge. Faculty members' supervision or guidance of students in recognized academic
pursuits that do not confer any University credit also is considered teaching. Associated
teaching activities include class preparation; grading; laboratory or equipment
maintenance; preparation and funding of proposals to improve instruction; attendance at
workshops on teaching improvement; and planning of curricula and courses of study.
Outreach/engagement activities such as service learning, conducting workshops, seminars,
and consultations, and the preparation of educational materials for those purposes, may be
integrated into teaching efforts. These outreach activities include teaching efforts of
faculty members with Extension appointments.

Excellent teachers are characterized by their command of subject matter; logical
organization and presentation of course material; formation of interrelationship among
fields of knowledge; energy and enthusiasm; availability to help students outside of class;
encouragement of curiosity, creativity, and critical thought; engagement of students in the
learning process; use of clear grading criteria; and respectful responses to student
questions and ideas.

Departments shall foster a culture that values and recognizes excellent teaching, and
encourages reflective self-assessment. To that end, departmental codes should, within the
context of their disciplines, (1) define effective teaching and (2) describe the process and
criteria for evaluating teaching effectiveness. Evaluation of teaching should be designed to
highlight strengths, identify deficiencies, and improve teaching and learning.

Evaluation criteria of teaching can include, but are not limited to, quality of curriculum
design; quality of instructional materials; achievement of student learning outcomes; and
effectiveness at presenting information, managing class sessions, encouraging student
engagement and critical thinking, and responding to student work. Evaluation of teaching

shall involve multiple sources of information such as course syllabi; signed peer evaluations; examples of course improvements; development of new courses and teaching techniques; integration of service learning; appropriate course surveys of teaching; letters, electronic mail messages, and/or other forms of written comments from current and/or former students; and evidence of the use of active and/or experiential learning, student learning achievement, professional development related to teaching and learning, and assessments from conference/workshop attendees. Anonymous letters or comments shall not be used to evaluate teaching, except with the consent of the instructor or as authorized in a department's code. Evaluation of teaching effectiveness should take into account the physical and curricular context in which teaching occurs (e.g., face-to-face and online settings; lower-division, upper-division, and graduate courses), established content standards and expectations, and the faculty member's teaching assignments, in particular the type and level of courses taught. The University provides resources to support the evaluation of teaching effectiveness, such as systems to create and assess teaching portfolios, access to exemplary teaching portfolios, and professional development programs focusing on teaching and learning.

Effective advising of students, at both the undergraduate and graduate levels, is a vital part of the teaching/learning process. Advising activities include, but are not limited to, meeting with students to explain graduation requirements; giving academic advice; giving career advice or referring the student to the appropriate person for that advice; and supervision of or assistance with graduate student theses/dissertations/projects. Advising is characterized by being available to students, keeping appointments, providing accurate and appropriate advice, and providing knowledgeable guidance. Evaluation of advising effectiveness can be based upon signed evaluations from current and/or former students, faculty members, and professional peers. The faculty in each academic unit shall develop specific criteria and standards for evaluation and methods for evaluating teaching advising effectiveness and shall evaluate advising as part of annual and periodic comprehensive reviews. These criteria, standards, and methods shall be incorporated into departmental codes.

FINAL INTERVIEW QUESTIONS DEVELOPED FOR [UNIVERSITY] FACULTY AND ADMINISTRATORS

- What excites you about this quality initiative to measure teacher effectiveness in [department]? [Ice breaker/intro]

1. The university code has a long list of what defines teaching and provides good examples of how it is portrayed – I'm interested in how you know when someone is a good teacher? [Define]

2. Again, referencing the university code and examples provided, what do you believe a faculty member has to do to be excellent? How would you suggest we quantify it? [Measure]

3. How should universities/administration measure teaching excellence? [Measure}

o What would be an excellent example for [department] to reference when assessing teacher excellence/what quality techniques should we seriously consider in measuring teaching excellence? [Define and measure]

4. In your experiences, what do you feel motivates faculty members who are excellent teachers? [Define]

5. What do you feel separates great teachers from their peers in reaching out to students? [Define and Measure]

6. If we are going to measure teaching excellence and we want to reward teaching excellence, can you provide examples of other organizations that have displayed a good reward system for teaching excellence? [Reward]

7. What do you see as the challenges we are going to encounter in implementing (defining and measuring) an excellence in teaching initiative beyond our department?

Request about opinion on publications – in discipline or education – equal weight?

8. Do you have any other thoughts that would help me and our advisory committee with this initiative?

APPENDIX C

FINALIZED FACULTY INSTRUMENT TO DEFINE TEACHING EXCELLENCE

Categories			
			For each category of statements, select the stated number of indicators detailed in blue by placing an **X** next to the quantity of indicators you feel **best illustrate**s teaching excellence for that particular category.

Academic Challenge	Choose 1		
			Student perseverance to grasp course content
			Challenge to students' prior understanding/paradigm of a concept or idea
			Course activity reflected the appropriate effort for the course level (i.e., 200-level, 400-level)

Student appreciation of the subject	Choose 2		
			Student desire to learn more about the subject
			Recommendation of the course by students to other students
			Relevance to students' future
			Relevance to real world issues
			Student enjoyment of the subject

Assessment of students	Choose 3		
			Assessment of student knowledge at beginning and end of semester (e.g., pre/post)
			Linkage between assessments and course objectives
			Rubrics developed and provided for each assignment
			Assignments provided options for original or creative work
			Assignments returned within 7 days
			Constructive feedback on assignments provided
			Multiple opportunities within the semester for students to give course feedback
			Assignments of grades aligned with stated criteria

Course Organization & Pedagogy	Choose 4		
			Allocation of class time for students to practice new skill(s) or technique(s)
			On-time start to class sessions
			Delivery of classes in an organized fashion
			Statement of course learning objectives in the syllabus
			Use of a variety of instructional methods
			Instructor enthusiasm for course content
			Summary of main points presented at the end of each class

		Effort of instructor to ensure student learning
		Integration of content from other disciplines
		Logical and sequential presentation of course content
		Update of course content/materials from previous years
		Presentations of real world examples in class

Student Learning	Choose 2	Students demonstrated analysis and synthesis of material
		Students demonstrated acquisition of facts and knowledge
		Student explanation about why key concepts are important
		Student application of learning to situations / life outside of school
		Increase in student understanding of the subject

Student Engagement	Choose 2	Students brought in examples about a topic after a topic was presented
		Classroom climate of respectful consideration for differing opinions
		Frequency of students questions, discussion and similar forms of participation
		Student attendance in class
		Attentiveness by students in class
		Student preparation for class

Community	Choose 3	Opportunities for group work
		Groups formed with varying levels of student competence and ability
		Students borrowed/ lent resources to each other (e.g. books, notes, etc.)
		Students interacted with each other to help better understand content
		Atmosphere of trust in class
		Course feedback periodically requested by instructor
		Improvements or change initiated based on student feedback
		Students encouraged to actively help create solution(s) for issues identified in class

Student/ Instructor Interaction	Choose 1	Sufficient access to instructor outside of class
		Encouragement by instructor for students to answer difficult questions
		Wrong answers responded to constructively
		Error or insufficient knowledge by instructor admitted, when applicable

SPRING 2015 COURSE EVALUATION PILOT TEST INSTRUMENT QUESTIONS

1. Class sessions increased my understanding of the subject.

2. I can see how class topics are relevant to issues in today's world.

3. I am interested in learning more about some of the topics in this class.

4. The teacher used a variety of instruction methods
   (e.g., group work, lecture, discussions, activities).

5. Helpful feedback was provided on assignments.

6. Time was allocated during class to apply new knowledge or skills.

7. The instructor periodically reviewed main points and concepts.

8. The instructor made an effort to ensure I understood course content.

9. The instructor used real life examples.

10. The instructor created an atmosphere that was respectful.

11. I came to class prepared
    (e.g., read assignments, watched videos, completed homework).

12. There were instances during this class when the course had all of my attention.

13. I was given opportunities to contribute in class.

14. I felt comfortable contributing in class.

15. I was given opportunities to provide feedback about the course.

16. The instructor was responsive to helping students.

17. Select the option that best describes your situation:

- ○ It is a required course
- ○ It was one of a list of courses from which I could choose to fulfill a requirement.
- ○ It is an elective course

18. I am a:

- ○ Freshmen
- ○ Sophomore
- ○ Junior
- ○ Senior
- ○ Graduate Student
- ○ Other

19. Gender

- ○ Male
- ○ Female
- ○ Prefer not to say

20. I expect to earn the following grade for this course:

- ○ A (+ or -)
- ○ B (+ or -)
- ○ C (+ or -)
- ○ D
- ○ F
- ○ Other

21. The course you are evaluating is: _____

WRITTEN COMMENTS. Please write your comments in the blank section below.

Please feel free to comment on any aspect of the **survey** (e.g., **question wording, number of questions, and clarity of questions**).

Your feedback will help to provide additional insight into improvements of **the survey**.

APPENDIX E

FACULTY MEMBER WRITTEN COMMENT RELEASE LETTER

Dear [department} faculty and instructors,

During the Fall 2014/Spring 2015 terms, a Task Force on Teaching Effectiveness assessment

was conducted. The project was conducted in HDNR on Defining and Assessing Teaching

Effectiveness. During that time, we gained access to the student course survey records from Fall

2011 through Spring 2015. These totaled ~39,000 records from the Warner College of Natural

Resources and were used as part of the Task Force project. These records provided one data

input for answering our questions on how do you define and assess teaching effectiveness in

higher education.

I have decided to use my experience and the learnings as a CSU staff member as my dissertation

topic: Defining and Assessing Teaching Effectiveness: A Case Study. I would like to obtain your

permission to use the records for research purposes in my dissertation.

The data I am requesting your approval on from Fall 2011 through Spring 2015 to use for

research purposes is your student course survey data records – qualitative data and Likert item

responses. I will maintain anonymity on these records by not using any specific identifiers of

you, other faculty, or your department.

The instructor names, course details, or personal identifiers will not be published. When reported

and shared, the data will be combined from all responses. While there are no direct benefits to

any of you, the goal of my research is to contribute to the knowledge on the consistency of students in providing feedback.

Thus, I need your written consent to be able to publish findings from these records. Your consent is voluntary and requires no additional time (other than a response to this e-mail). If you decide to provide consent, you may withdraw your consent at any time without penalty.

There are no anticipated risks to you from providing consent. It is not possible to identify all potential risks in research procedures, but the researcher(s) has taken reasonable safeguards to minimize any known and potential (but unknown) risks. As we have approached the end of another semester, please respond to this e-mail with your consent. If I don't hear back from you by Friday, December 1st, 2017, I will re-contact you.

If you have any questions about the research, please contact Michele Marquitz at Michele.marquitz@colostate.edu or Carole Makela at carole.makela@colostate.edu. If you have any questions about your rights as a volunteer in this research, contact the CSU IRB at: RICRO_IRB@mail.colostate.edu; 970-491-1553.

Michele Marquitz

PhD Candidate

APPENDIX F

INSTITUTIONAL REVIEW BOARD APPROVAL

Colorado State University
Knowledge to Go Places

Research Integrity & Compliance Review Office
Office of the Vice President for Research
321 General Services Building - Campus Delivery 2011 eprotocol
TEL: (970) 491-1553
FAX: (970) 491-2293

NOTICE OF APPROVAL FOR HUMAN RESEARCH

DATE:	May 02, 2017
TO:	Makela, Carole, School of Education
	Jennings, Louise, School of Education; Marquis, Michele, School of Education
FROM:	Swiss, Evelyn, CSU IRB 2
PROTOCOL TITLE:	Defining and Assessing Teaching Effectiveness in Higher Education: A Case Study
FUNDING SOURCE:	NONE
PROTOCOL NUMBER:	15-5739H
APPROVAL PERIOD:	Approval Date: May 02, 2017 Expiration Date: April 27, 2018

The CSU Institutional Review Board (IRB) for the protection of human subjects has reviewed the protocol entitled: Defining and Assessing Teaching Effectiveness in Higher Education: A Case Study. The project has been approved for the procedures and subjects described in the protocol. This protocol must be reviewed for renewal on a yearly basis for as long as the research remains active. Should the protocol not be renewed before expiration, all activities must cease until the protocol has been re-reviewed.

Important Reminder: If you will consent your participants with a signed consent document, it is your responsibility to use the consent form that has been finalized and uploaded into the consent section of eProtocol by the IRB coordinators. Failure to use the finalized consent form available to you in eProtocol is a reportable protocol violation.

If approval did not accompany a proposal when it was submitted to a sponsor, it is the PI's responsibility to provide the sponsor with the approval notice.

This approval is issued under Colorado State University's Federal Wide Assurance 00000647 with the Office for Human Research Protections (OHRP). If you have any questions regarding your obligations under CSU's Assurance, please do not hesitate to contact us.

Please direct any questions about the IRB's actions on this project to:

IRB Office - (970) 491-1553, RICRO_IRB@mail.Colostate.edu
Evelyn Swiss, Senior IRB Coordinator - (970) 491-1381; Evelyn.Swiss@Colostate.edu
Tammy Felton-Noyle, Assistant IRB Coordinator - (970) 491-1655; Tammy.Felton-Noyle@Colostate.edu

Swiss, Evelyn

Initial Approval has been granted to use existing data collected for nonresearch purposes with the approved recruitment and consent procedures. The above-referenced research activity has been reviewed and approved by the Institutional Review Board under expedited review category 5. Approved documents include: Faculty Consent Cover Letter; Department Had Consent Cover Letter; Chapter 3 IRB Document Submittal 1; Protocol ID Cycle 2,

Page 1

159

Colorado State University

Knowledge to Go Places

Research Integrity & Compliance Review Office
Office of the Vice President for Research
321 General Services Building – Campus Delivery 2011 sprotocol
TEL: (970) 491-1553
FAX: (970) 491-2293

04152017

Approval Period:	May 02, 2017 through April 27, 2018
Review Type:	EXPEDITED
IRB Number:	00000202