2015-03-01

Student Voices in Teacher Evaluations

Brenda Sue Burr
Brigham Young University - Provo

Student Voices in Teacher Evaluations

Brenda Sue Burr

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctorate of Education

Julie Hite, Chair
Legrand Richards
Scott Ferrin
E. Vance Randall
Sterling Hilton

Department of Educational Leadership and Foundations

Brigham Young University

March 2015

ABSTRACT

Student Voices in Teacher Evaluations

Brenda Sue Burr
Department of Educational Leadership and Foundations, BYU
Doctor of Education

In an ever increasingly competitive global marketplace, a concern exists that American students are not being adequately prepared with the skills needed for the 21st century. As a remedy, improving quality of teacher instruction is a current national focus. Stakeholders are questioning current infrequent and inefficient methods of evaluating teacher performance. Many states are looking at using a 360 model of evaluating through multiple perspectives including the students themselves as key stakeholders.

One method of accessing student voice and adding another perspective to teacher evaluations would be to include student evaluations in the rating of teacher performance, Student Evaluation of Teaching (SET). While using student evaluations of teacher performance is widespread in higher education, the practice has been limited in public school settings until brought to light by the publication of the recent Gates Foundation MET (Measures of Effective Teaching) Project (2010). Currently, states across the nation are considering adding a student input component to teacher evaluations. With the validity and reliability of student evaluations in the university settings still under debate by professors, public school teachers also fear punitive measures and public judgment based on the verdicts of adolescents.

This research examined the archival data from a program study of one high school's student evaluation implementation process, accessing teacher feedback from the initial evaluation process and then an adjusted implementation of student evaluations according to teacher feedback the following year. Based on mixed method design using both qualitative and quantitative methods to analyze teacher questionnaires, focus group open-ended responses and statistical analysis of close-ended agree/disagree statements from teacher questionnaires, this study used triangulation to explore teacher reflections on their anxiety levels created by the student evaluation implementation process, the value they found in student evaluations, and the degree to which student evaluations facilitated change in their teaching instruction. Exploring possibilities through the eyes of teachers to reduce their anxiety and increase their value of student input, this study suggests ways to tap into the potential but underutilized resource in schools that could come from developing a mutually beneficial partnership between students and teachers to improve teacher instruction and increase student learning.

Keywords: 21st century skills, teacher evaluation, 360 method, student voice, stakeholders, student evaluation of teaching (SET)

ACKNOWLEDGEMENTS

I wish to express my gratitude to my committee chair, Julie Hite, who never failed to give timely advice, meticulous scrutiny, thoughtful insights, and on the spot expert instruction. I would also like to thank my entire doctoral committee, Legrand Richards, Vance Randall, Scott Ferrin, and Sterling Hilton for their support. A special thanks to Cliff Mayes, for his encouragement and confidence in me that inspired me to begin this program and to persevere. To my cohort, my friends, I owe a deep debt of gratitude. We shared the triumphs and trials of the program and much of life together the past three years.

For my educational roots, I must thank my dear mother. "Richer than I you could never be. I had a mother who read to me." She is still reading at 82, the prime example of a life-long learner. She is my biggest fan and my heroine of faith and unconditional love.

To my own children, who have endured many a year with a mother who was insistent on them doing their homework, who was consistently doing her own, and who was constantly grading the homework of others, I express my love and gratitude. "School stories," as they are affectionately referred to in our home, were faithfully endured as often they willingly sacrificed their mother's time for the sake of a troubled student who needed her more. I am proud of their love of education and their chosen work. They are and always will be, my world and my life's greatest work.

And finally, to my dear eternal companion, I express my deepest love and gratitude. The road has been long and often rough and filled with many a detour, but holding hands, the young couple who began 36 years ago with a dream and one college semester a piece will finally be Dr. and Dr. Burr.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Background

Increasing measures of accountability for schools, teachers, and students are at the heart of educational reform for American students. One of the new accountability measures being implemented is a reformation of the evaluation of teacher performance. The implementation of a new multi-perspective or multi-stakeholder evaluation process that includes the voice of the students is being considered in many school districts across the nation. While the use of student evaluations in the university setting is common but controversial, the debate now extends to their use in public secondary and elementary schools. This research will focus on whether the concerns over the reliability and validity of student evaluations and teacher anxiety over their potential punitive application can be overcome and the value of formative feedback and the inclusion of student voice heightened to create a positive student-teacher partnership that increases the performance of both.

Teacher Quality in American Schools

From the mid 1980's to the mid 1990's democratic leadership was heralded as the leadership style needed to increase teacher productivity and morale. It also sought to change traditional schools into communities of stakeholders that fostered increased learning and active participation using leaders willing to question traditional bureaucratic practices and the outdated top down framework of schools (James & Rottam, 2009).

During this time, business took an interest in education policy and discussions centered on a teacher quality agenda. The Carnegie Task Force on Teaching as a Profession released *A Nation Prepared: Teachers for the 21st Century* in 1986 that called for the creation of National Teaching Standards. In the same year the business-led Committee for Economic Development

(CED) issued *Investing in Our Children* (Koppich & Esch, 2012) which recommended "nothing less than a revolution in the role of the teacher in order to upgrade the quality and professionalism of the U.S. teacher workforce" (CED, 1985, p. 60).

As the economy of the late 1990's improved, business interest in education began to slow down and policy making shifted more to federal reforms. The first ever national educational standards for teachers were developed during the Bush Administration. With the passage of the No Child Left Behind Act (NCLB) in 2001, the federal government held states responsible for the achievement level of all students and mandated that schools have *highly qualified teachers* (Koppich & Esch, 2012, p. 85). At the same time philanthropies, such as the Gates, Ford, Joyce, Dell, Broad, Walton, and Milken foundations came to be an influence on educational policy (DeMarrais & Suggs, 2011).

The urgency of school reform now has reached a feverous pitch. According to the 2013 Program for International Student Assessment (PISA) scores, international comparisons of student achievement ranked the United States 26th in math, 21st in science and in 17th place for reading. While there is truth to the argument that these tests are not comparing "apples to apples" with many differing factors among countries included who is tested, the results still show the United States is actually the exception rather than the norm staying at about the same level for decades as other countries passed up the US (Ripley, 2013), with some showing remarkable improvement in 2013, for example, Viet Nam (PISA, 2013). American schools are under scrutiny to increase student achievement and regain a competitive advantage in the global marketplace.

Teacher Evaluations

The largest of the new federal education programs under the Obama administration was *Race to the Top*. The $4.35 billion in federal stimulus funds was offered to states with progressive education reform plans and a commitment to teacher effectiveness (Koppich & Esch, 2012). Forty states and the District of Columbia changed state laws to become eligible for Race to the Top funds including the development of new state policies that focused on changing the ways teachers are evaluated (Zinth, 2010). On February 9, 2012, Secretary of Education Arnie Duncan launched a five million dollar proposal entitled the RESPECT Project (Recognizing Educational Success, Professional Excellence, and Collaborative Teaching) calling for teacher evaluations based on the 360 method or multiple perspectives and multiple stakeholders (U. S. Department of Education, 2012).

The *360 method* of multiple stakeholders evaluating performance was developed by businesses. Team evaluation or 360 evaluation meant that an employee would be evaluated by all who have contact with them including supervisors, peers, clients, and the public or a client-driven evaluation system. In an educational setting, the 360 method of feedback included feedback from principals, peers, parents, and students, as well as self-reflection and assessment of student achievement (Manatt, 1997). Policy makers across the nation are proposing legislation that requires school districts to establish new systems of teacher evaluation that commonly require multiple measures of performance, including classroom observations, student scores and measures of student growth on standardized tests, and surveys of parents and students (Strunk, 2012).

New, more demanding, multiple perspective evaluations of teacher effectiveness as measured by student outcomes and student input were also the focus of the $45 million dollar

educational MET (Measures for Effective Teaching) research project of the Gates Foundation (2010) and serves as a new directive to school districts across the nation. Even a discussion around teacher evaluations is earth shattering for the education industry. Historically, teacher reviews have been haphazard, ranging from nonexistent to an annual classroom visit from the principal — often referred to as the drive-by or stop-buy (Smith, 2012).

Students as Teacher Evaluators

The MET (2010) project highlights students as valuable stakeholders in this multiple perspective process. "Students belong in a category by themselves. They are the most intimately involved with and aware of the school's needs and successes—at the same time they are the least integrated into analysis, decision-making, and planning processes" (Holcomb, 2004, p. 39). Fullan (2001) also noted:

> When adults think of students, they think of them as beneficiaries of change. They rarely think of students as participants in a process of change. Too little has actually happened to enhance the role of students as members of the school as an organization. . . . Unless they have some meaningful role in the enterprise, most educational change, indeed most education, will fail. (p. 151)

Student voice. Student voice can be defined as the many ways in which students might be given the opportunity to participate in school decisions that will shape their lives and the lives of their peers (Fielding, 2004). Community unrest over inequality sparked and spilled over into student voice movements in the 1960s and 1970s where high school students began to assert the right to participate in decision making, and to be included in the practice of democracy by having a say in their education. Yet a focus on the actual role or implementation of students in school decision making practices and teacher evaluations largely fizzled out (Levin, 2000).

Research is still sparse on just how to facilitate the development of student voice and promote youth leadership which embodies actual meaningful roles and responsibilities (Mitra, 2005) within school reform and teacher evaluations. Thus, this research explores and investigates possibilities to make students among the valued stakeholders, giving students an authentic voice in improving their education and giving them a say in assessing teacher performance.

Teacher fears. As evidenced by the recent teacher strikes in Chicago, some teachers fear new evaluation measures (Pearson, 2012). They are afraid that formulas will put the blame solely on teachers for poor student performance when so many complicated factors come into play. Teachers agree that the quality of instruction matters, but they question the methods used to quantify quality. The issue becomes even more contentious with the possibility of teacher evaluation scores being linked to pay and to continued job security (Smith, 2012). Teachers fear evaluations would turn into popularity contests (Zabaleta, 2007) in students' eyes and question how decisions that might determine job security and instructional practices could be left to the whim of and maturity of adolescent opinions.

Teachers are definitely concerned over the validity, bias, and reliability issues surrounding student evaluations, Teachers also question how student evaluations will be used in determining job security. Can these teacher concerns be alleviated while still using student evaluations as a way to inform and improve teacher instruction and access student voice? This is the question this research seeks to answer.

National Focus on Teacher Evaluations

Despite teacher fears over new evaluation methods, The National Education Association (NEA), the nation's largest public school employee organization, recently released an official

statement regarding teacher evaluations (Policy Statement on Teacher Evaluation and Accountability, 2011).

> NEA believes that our students and teachers deserve high quality evaluation systems that provide the tools teachers need to continuously tailor instruction, enhance practice and advance student learning. Such systems must provide both ongoing, non-evaluative, formative feedback and regular, comprehensive, meaningful and fair evaluations. Such systems must be developed and implemented with teachers and their representatives, either through collective bargaining where available, or in partnership with the affiliate representing teachers at the state and local level.

This policy statement was seen as heralding the possibility for greater partnering and cooperation between teacher union leaders with local public school officials. NEA President Dennis Van Roekel "believes the new statement signals a commitment to a new, more prestigious profession of teaching and reflects the first broad endorsement by NEA of the need for evaluation and accountability reform" and calls for "robust evaluations based on multiple indicators" (Walker, 2011, p. 1). While embracing reform in the evaluation of teachers, Van Roekel also expressed concern that the implementation process should involve a collaborative effort including teacher input. "As more states and districts seek to improve teacher evaluation, the risk is that reform is done to teachers rather than with them" (Walker, 2011, p. 1).

Maryland model. The Montgomery County Education Association in Maryland is one example of teacher associations partnering with school systems and legislatures to revise teacher evaluations. The MCEA played a vital role in the creation and implementation of a new teacher evaluation policy that includes multiple perspectives and that no longer pitted teachers against each other for merit pay (Sullivan, 2012). Sullivan explained the teachers' sentiments.

> This bothered many teachers in the MCEA who saw teaching as a cooperative profession, with teachers working together for the common cause of educating every child. . . Comparing teachers to one another and determining who was the most outstanding seemed to encourage teachers to see other teachers as their competition, not their colleagues. . . Teachers, as part of a team of educators, appreciate that everyone benefits when every teacher in a school is doing his or her job well (pp. 142–143).

The pathway to partnering teachers and other stakeholders in effectively improving education is a crucial issue. "Unpacking the elements of effective classroom instruction—discovering the policies and practices that maximize teachers' ability to boost levels of student achievement—is key to improving American education" (Koppich & Esch, 2012, p. 80).

Lee County model. As part of a school reform plan entitled *Choosing Excellence*, The collaborative efforts of the School District of Lee County, in Lee County, Florida, included the Foundation for Lee County Public Schools and the Teachers Association of Lee County in systematically requiring teachers to elicit student feedback to aide in teacher goal setting and to refine instructional practices irrespective of whether or not the feedback would be tied to teacher evaluation systems. Mark Castellano, president of the Teachers Association of Lee County stated that

> As advocates for the professionalization of teaching, our union embraces data and information that teachers can use to become more effective and accountable teachers. We are committed to our students, to their learning. Student feedback in *Choosing Excellence* schools has created a new bridge between teachers and their students, more of a dialogue really, about what appears to be working and what is not (Sanford, 2013, p. 6).

Utah education legislation. Other states are in the process of changing their teacher evaluations. One of those states is Utah. Utah Legislative House Bill 64, *Public Education Employment Reform* (2012), outlines stipulations for new more rigorous evaluations of teacher performance. "The Legislature recognizes that the quality of public education can be improved and enhanced by systematic, fair, and competent annual evaluation of public educators" (lines 347-349) and requires that those evaluations include "multiple lines of evidences" (line 406) of which one possible source is "student input" (line 408) and that salary increases or the "compensation system to be aligned with the district's annual evaluation system" (lines 665-666). The development of methods to facilitate these new teacher evaluation requirements by the 2014-2015 school year are currently in progress in Utah.

Many other states across the nation are also exploring new methods for teacher evaluations (Strunk, 2012). The Utah bill allows a local school board to develop its own evaluation program, within guidelines set by the State Board of Education, or adopts an evaluation program developed by the State Board. As long as all requirements of the bill are met, there can be local control over implementation. The specifics of how the results of teacher evaluations will be used in salary compensation measures, how results will be published, and what percentages will be given to each component or perspective in the new teacher evaluation plan remains uncertain. These measures are still in the developmental stage by the State School Board and Utah State Office of Education. The creation of evaluation tools and the design of the implementation at the state level are being developed in a collaborative process with school districts partnering with the state office of education in pilot programs.

Research Purpose

"Few measures of teachers' classroom ability inspire as much optimism among researchers—and as much unease among educators—as surveys of students" (Cavanagh, 2014, p. 1). With the 21st century paradigm shift "from an instructional paradigm (one that puts the teacher and instruction at the center of teaching) to a learning paradigm (one that places the learner at the center of teaching)" (Calkins & Micari, 2010, p. 14), comes the opportunity for new ways of thinking about student evaluations. This shift provides the opportunity for "questions turned from whether and how evaluation tools may or may not be telling the truth to how evaluation can be conducted in more meaningful ways" (p. 15). Chulkov and Van Alstine (2012) called for the continuous review and improvement of the student evaluation process.

This research investigates the potential and possibilities for creating an implementation process for student evaluations that would unite teachers and students, seating them together around the stakeholder table. The researcher will study the archival data from the implementation of two rounds of student evaluations specifically through the eyes of the teachers receiving the feedback at a high school in Utah.

The researcher was a member of the Utah State Teacher Effectiveness Committee whose purpose was to give input to the design of the principal's new observation evaluation tool. The uncertainty over how House Bill 64 will be implemented is causing anxiety. Many districts are preparing to look at that part of the process for which there is local control and to prepare for implementation of state mandates.

This study seeks to examine the student evaluation implementation process at one Utah high school by analyzing the initial implementation of student evaluations (Phase I) and a re-implementation the following year (Phase II) after modifications suggested from the initial

implementation. The teachers gave feedback on the implementation process of both phases through teacher questionnaires and focus group discussions. The researcher will access and analyze this feedback from both phases.

Research Questions

At this important turning point, "states must not respond with quick-fix measures to address teacher quality by simply imposing rigorous evaluation standards, which will result in the disaffection of teaching professionals" (Oon-Seng, 2012, p. 76). Strunk (2012) warns against the rush to implement new multiple perspective teacher evaluation policies which could undermine their potential benefits, advocating instead allowing time to learn from initial implementations in a no- or low-stakes setting. This research study examines one implementation study in just such a setting**.**

The research hopes to examine the overall question of how teachers react to SETs, shedding light on two main sub-research questions.

1. Will teachers find value in student voice?
    a. Will the value level of student evaluations be sufficient to cause teacher reflection?
    b. Will the value level of student evaluations be sufficient to impact teacher instruction?
    c. Does teacher anxiety level over student evaluations impact teacher value?
    d. Will teachers belief in the ability of students to evaluate impact value?
2. What can be done to improve the implementation process of giving students voice through SETs?

a. Will teachers believe that shared ownership and a collaborative process improves the process?

b. How can the implementation process be improved to increase teacher value?

This research involved teachers in analyzing and determining the value of student evaluations and seeks for improving the process to allow for non-threatening ways for teachers to focus on examining their instruction from a student perspective. The methods in this study include both qualitative and quantitative analyses, using a mixed method design, of teacher responses through questionnaires and focus group discussions. The research used archival data from two phases, one year apart, of the implementation of school-wide student evaluations of teacher performance at one Utah high school. A third non-archival phase was implemented based on the findings of the first two phases.

Summary

Teaching is personal; to be done well, it requires a piece of the heart. Thus, being the target of an evaluation of a very personal performance, it is hard to avoid not taking it personally. Cohen (1990) summed up the concerns: "Negative attitudes toward student ratings are especially resistant to change." Teachers "support their belief in student rating myths with personal and anecdotal evidence, which for them outweighs empirically based research evidence" (pp. 124–125). Perhaps, if the argument can be changed from merely examining the accuracy or inaccuracy of student evaluations to rather examining or selecting feedback for helpful insight, students and teachers could work together to improve instruction. As John Daley (1999) stated in an article in the *NEA Higher Education Journal*, if administered fairly, "I believe enough in the student evaluation's diagnostic potential to risk the impertinent and the downright nasty in order

to glean a few useful insights" (p. 57). It is the possible worth or value of those insights and the outcome that is the focus of this research.

If teacher concerns can be resolved over student evaluations, the discussion of reform, as the MET project suggested, perhaps, could begin to tap in to the potential but underutilized source of power that could come from developing a partnership with the students themselves. "Educational reformers often partner with others to make change happen in their schools. But few reformers look to students as agents of change" (Yonezawa & Jones, 2009, p. 205). Perhaps, increasing school performance means including the ones performing in the power, both teachers and students. Gaining or regaining the competitive edge internationally may include giving American students and teachers together the chance to fully participate in education reform.

Chapter 2

Review of Literature

This chapter will cover topics related to both the value of accessing student voice as well as the controversy over how that access is to be achieved. The review refers specifically to the current reform movement in education calling for a new multiple perspective method of evaluating teacher performance and the potential positive use of student voice in the form of student evaluations as one of those perspectives.

Student Voice

Student voice is the unique educational perspective of the students. When that voice is included, students become actual stakeholders investing in their own education. Important concepts about student voice are found in examining student voice as it relates to current practices in education, connections with student achievement, and opportunities for increasing student voice.

Historical perspective. Leadership within a school setting can encompass the idea that people through their leaders "ought to be able to shape the institutions, culture and relationships of which they are a part" (James & Rottman, 2009, p. 478). Embracing Professional Learning Communities (PLC), many schools have changed to a much more collaborative environment than the previous culture of isolation (DuFour & Eaker, 1998). Administrators empower teachers as they guide school improvement rather than mandate. Administrators derive their power from empowered teachers who are much more likely to follow a leader who implements policies in which they had a say. "When schools move into sharing of authority, collective identities, communities of practice, and serving others, a more democratic learning community emerges"

(Williams, Cate & O'Hair, 2009, p. 458). Thus, to achieve the full collective identity, a school must consider including as many stakeholders as possible, including the students.

Taking a second look at the relationships that administrators and teachers build with students, the very people they serve may be the very source of new power to facilitate improvement in student achievement.

Current practice. Although the idea of allowing for student voice is not new to education, it has experienced resurgence in recent years. And yet, in the idea of leaving no child behind came the emphasis on accountability for student outcomes and a focus on measuring the success of those outcomes by test scores while the broader mission of preparing students to become engaged and contributory citizens moved back to limbo. Increased demand for accountability and visible results of student achievement has narrowed the vision and purpose of schooling in recent years, not only in terms of pedagogy and content, but also participation. "As the accountability movement has been designed and implemented with little student input, one must question its ability to increase engagement of high school students" (Mitra & Gross, 2009, p. 525).

However, No Child Left Behind has been replaced by the reality of not leaving America behind. "Developing countries like China and India are offer[ing] the world economy workers of increasing education and sophistication at far lower costs than the USA can match" (Center for American Progress, 2005). Secretary of Education, Arnie Duncan, was quoted in the New York Times (Dillon, 2010) on the release of the 2009 international educational scores.

> We have to see this as a wake-up call. I know skeptics will want to argue with the results, but we consider them to be accurate and reliable, and we have to see them as a challenge

> to get better. The United States came in 23rd or 24th in most subjects. We can quibble, or we can face the brutal truth that we're being out-educated (p. 1).

On March 10, 2009 in a speech on education delivered to the U.S. Hispanic Chamber of Commerce, President Obama (2009) gave this now often repeated quote: "Despite resources that are unmatched anywhere in the world, we have let our grades slip, our schools crumble, our teacher quality fall short, and other nations outpace us. . . . The future belongs to the nation that best educates its citizen" (p. 1). Concerns over US students being less prepared than their international cohorts has led to a rigorous remedy, a new Common Core. The Common Core focuses on 21st century skills most needed for future career success (USOE, 2014). The four Cs as they are called are critical thinking, collaboration, creativity, and communication. "The educational mandate has shifted from ensuring that *all* students learn to ensuring that *all* students learn at *high* levels" (Dufour & Marzano, 2011, p. 141). With a push from the new Common Core to facilitate higher levels of thinking including the ability to "evaluate intricate arguments; and the capacity to surmount the challenges posed by complex texts, create and test solutions, and exert opinions in a rigorous classroom discourse" (USOE, 2014), it seems contradictory to not give students more consideration in making decisions about their own education.

In the educationally and culturally diverse setting of American, The Center for American Progress and the Institute of America's future (2005) calls for the American education system to rise to the challenge facing our country and the educational process needed to continue to ensure education that supports democracy by stating

> We must ensure that all American children—regardless of race, ethnicity, income, native language, or geographic location—are afforded access to high-quality schools that will enable them to participate in the promised opportunity of the American dream. Failure to

> do so will only lead to greater divisions in the country between the haves and have-nots, which history tells us can have disastrous consequences. We must also produce more high-caliber students to compete successfully with young people overseas (p. iii).

With the urgency of school reform and with a need for greater focus on higher levels of learning for all students being the goal, Smyth (2006) questions the effectiveness of any program that does not first address the simultaneous increased disengagement of students with schools and learning. "We need a more mature and nuanced approach that is more inclusive of those most affected, and by that I mean students" (p. 288). In other words, the problem must be looked at through the positional lens (Glazier, 2005) of those being served, the students. Fullan (2001) notes that

> When adults think of students, they think of them as beneficiaries of change. They rarely think of students as participants in a process of change. Too little has actually happened to enhance the role of students as members of the school as an organization. . . . Unless they have some meaningful role in the enterprise, most educational change, indeed most education, will fail (p. 151).

Research suggests then that students should be respected as expert witnesses by reconfiguring "the power dynamics and discourse practices within consisting realms of conversation about education to embrace the student perspective" (Cook-Sather, 2002, p. 3; Rudduck & Demetriou, 2003, p. 290), "such collaborative engagement causes students to feel respected and this often produces higher levels of student involvement in both their learning and the school" (Cook-Sather, 2006, p. 360). Actually seeking out and listening to the perspectives of students can serve as a method for quality control, and unlike so many other reform strategies- this one costs little to nothing. Given the importance of what is at stake in our efforts to reform

the nation's schools, it may be time to try an approach that allows us to learn about how to improve schools without having to expend additional resources while also engaging those with so much at stake—the students (Nogurea, 2006).

Connection to student achievement. Anonymous student surveys addressing school climate issues and concerns are not new and have been part of school accreditation procedures. However, with the rise of issues of school safety, the U.S. Department of Education awarded eleven states a piece of a $37 million dollar grant to explore school climate through the eyes of multiple perspectives including students (Shah, 2013). David Osher, the principal investigator at the National Center on Safe Supportive Learning Environments said, "It's really understanding that what you want to do is create emotionally safe and supportive conditions in school so people work together better and learn better together" (p. 2). This same idea of working together with the students is spreading beyond the classroom climate to gathering student input on classroom instruction.

Perhaps education is overlooking the very key to increasing student achievement; increasing student connection to school by increasing their chance to have a say in the learning process. In considering the needs of adolescents for *connectedness* (Smyth & McInerney, 2006) along with the four C's of the common core, educators may be able to reach the hearts of students as well as their minds by letting them speak.

> Evidence suggests that an important first step in including students in the reform process is to give students a 'voice'—in other words, to give them the right to speak for themselves about their educational experiences and the opportunity to be heard by those in places of power (Feuer, 2009, p. 17).

This concept of student voice incorporates much more than the current idea of student leadership given to most students through student councils with power limited to planning assemblies and school dances. All students, from those making straight-A's to those struggling to pass, need to believe that they have a voice in their education and need to be given the chance to voice their ideas (Feuer, 2009).

When students feel that their lives, experiences, cultures, and aspirations are ignored, trivialized, or denigrated, they develop hostility to the institution of schooling (Feuer, 2009). Culture and climate surveys asking for student, parent, faculty, and staff input have proven remarkably valid. School climate measures were associated with student academic performance. In 2009 the U.S. Department of Education's Office of Safe and Drug Free Schools drafted a model of school climate that includes three main categories: engagement, safety, and environment. "High-quality relationships between and among teachers and students are critical for developing a school environment conducive to student attachment and learning (O'Malley, Renshaw, Ritchey, & Furlong, 2011).

The truth may be that while Wall Street senses the economic crisis and the future global marketplace calls for increased levels of learning, many students simply do not care because they are not invested, and therefore not achieving. *Student engagement* means more than involvement in activities or the lesson for that day, but engaged in the very mission of learning itself.

> By elevating student voice to its rightful status, we can change the way that students view learning, themselves, and their school. . . By listening to student voices, we can motivate and engage student's in today's schools, and that engagement can lead to greater achievement. . . Those critics who do not address low levels of student engagement link it to other factors, only rarely, considering student engagement in its own right. . . . Unless

> we engage our students—unless we get them to care—not much else will matter (Beaudoin, 2005, p. 5).

Not caring could be the way students are manifesting the need for control. By the very act of avoidance or disengaging, students may be making a statement, overtly or covertly or inadvertently by choosing not to learn. Choosing not to learn can be a form of political resistance (Erickson, 1987). Motivation research suggests that as humans we are more motivated when we have say. Maslow's hierarchy of human needs implies that the opportunity to exercise control or have some freedom in choices is an essential human need linked to motivation to act (Maslow, 1943).

In the eight-year study of American high schools written about by John Goodlad in *The Moral Dimensions of Teaching* (1990), he quoted Herbert Thelan, one of the participants in the study. Thelan concluded that by separating the "creating, planning, thinking, and doing" (p. 286) from the students themselves in the high school setting, education is in fact separating students from being able to attach a purpose or meaning to their education. Thelan continued:

> Since the practical importance of purpose is to enable us to see how to recognize and choose among alternatives, the practical consequence of avoiding purpose is avoidance of the necessity for choosing, and with this of course, the flight from freedom, for freedom without choice is impossible (p. 286).

The alternative would then be to make students into "powerful people" (Smyth, 2006, p. 292) by returning to them some control or say over their own learning and the learning environment.

In their research on using the student perspective in school reform, Bergmark and Kostenius (2009) found students not only capable and competent but also trustworthy in their

abilities to enrich the process of school improvement. "Increasing student voice in schools also has been shown to help to re-engage alienated students by providing them with a stronger sense of ownership in their schools" (Mitra, 2003, p. 289). It seems that crucial information on how to improve our nation's school system might reside with those who will benefit the most—the students themselves (Boser & Rosenthal, 2012).

Measuring student voice. Major researchers in student voice have developed tools to measure the level of authenticity of student participation. Roger Hart's (1992) worked for UNICEF as a child right activist. His ladder of student participation model describes the levels at which young people participate within organizations in the public domain. By addressing those people who have it in their power to give young people the opportunity to have a voice or higher levels of participation, Hart hoped to inform adults including school leaders who were unknowingly trivializing or only giving lip service to student involvement. His ladder starts with tokenism and manipulation, or non-participation, and moves to the highest level that involves student-initiated action and shared power with adults, eventually facilitating youth who grow up prepared to enter their communities as active participants (see Figure 1).

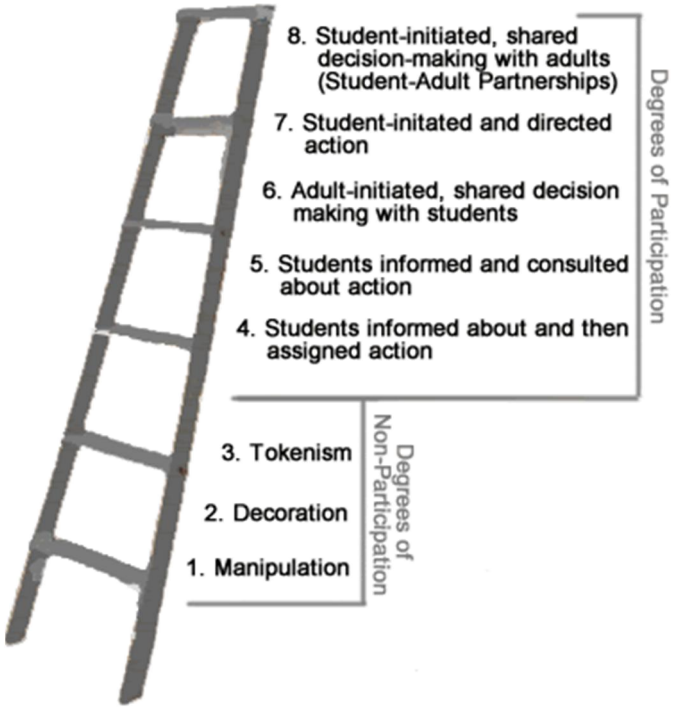


Figure 1. Hart's ladder of participation (1992, p. 8).

Michael Fielding's scale of student involvement (2001) proposed and practiced the idea of student researchers at a high school in the UK with "a transformative notion of education at the heart of which lies the commitment to teaching and learning as a genuinely shared responsibility" (p. 137). His student involvement scale (see Figure 2) begins with the student as a data source, active respondent, co-researcher, and finally as researchers who initiate change.

	Classroom	*Department / Team*	*School*
Students as **Data Source**	Past performance data	Samples of pupil / student work	Pupil attitude surveys
Students as **Active Respondent**	Shared lesson objectives	Dept / team agenda based on pupil perceptions	School council
Students as **Co-Researcher**	Feedback to teacher	Joint research on independent learning	Joint review of rewards system
Students as **Pupil / Student Researcher**	What Makes a Good Lesson?	Research + suggestions for new units of work	Pupil led review of tutorial programme

Figure 2. Scale of student involvement (Fielding, 2001, p. 137).

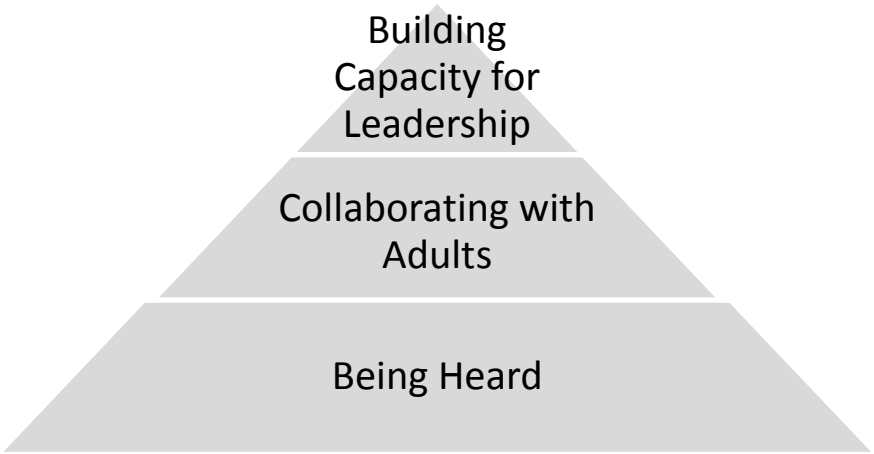


Figure 3. Pyramid of student voice (Mitra, 2006, p. 7).

Dana Mitra (2006), a former elementary teacher, Penn State professor, Fulbright Scholar, and student voice advocate, created a pyramid of student voice (see Figure 3) that illustrates the correlation between the increases in possibilities for youth development that occur as student voice increases. The pyramid starts at the lowest level with students being heard, moving to

collaborating with adults, and finally to building capacity for leadership. In initiating the stages of Mitra's pyramid at Whitman High School, the school leaders purposely focused on changing the student-teacher role to one of partnership and trust. One student reflected on his experience in a student forum. "When I was talking to those teachers, you could just see those eyes of people who just wanted to know what we were thinking. That just felt so powerful" (p. 10).

In all of these spectrums, the researchers assert that control must eventually go beyond the typical levels of just listening to the voice of students, but listening may be a place to begin. Letting students be heard can encourage school personnel to challenge their assumptions and begin to see alternative solutions and possibilities for partnering (Mitra, 2006).

Increasing student voice. Some states are including student membership on community or site based councils, and some districts include students as school board members or include a student or youth school board that reports to the local school board. In 2003, 28 of 35 states that responded to a survey by the National School Boards Association reported that students serve on local school boards throughout their states. In California, state law grants these student members full voting rights (Feuer, 2009).

While these steps represent a shift towards encouraging student voice, many of these students still serve as non-voting members and the students on the boards or councils typically are appointed from students already serving on student councils or in student leadership positions. By not including the voice of all students, and particularly those students who are failing subjects or rarely attending school, schools may find it easier to shift the blame for failure onto students rather than to look within at school culture and leadership (Mitra, 2005). Some schools are looking at expanding leadership opportunities to allow for students from diverse student populations, including all club presidents, and leaders representing diversity from groups

such as Latinos in Action, athletic team captains, and on student governing councils. One school created a Student Forum to address a school improvement plan with a "cross-section of the student population for the focus groups on the basis of race, gender, academic performance, and cliques" (Mitra, 2006, p. 7).

In developing and choosing new curriculum, schools have arranged for students to serve on curriculum committees. Schools have given students a role in staff development sessions or faculty meetings to give feedback on particular programs (Fielding, 2001). In fact, youth participation in faculty meetings has often changed the tenor of conversations, including reducing unprofessional behaviors such as completing crossword puzzles during staff meetings or displays of hostility to colleagues (Mitra, 2003).

Increasing student voice is also viewed as increasing choice. Many students now plan their schedules by registering on-line, choosing what class to take, what period to take it, and the teacher they prefer. Variety in course selection and pathways, on-line courses, distance learning, concurrent enrollment, and specialized charter schools are also expanding to give students more control over their educational path. Internships and work-based learning and other vocational training prospects are increasing the opportunities for students to participate in hands-on apprenticeship as an alternative method for learning. Service-learning opportunities are also increasing as many schools offer their students the opportunity to participate in tutoring programs and partnerships with local volunteer organizations. Going beyond choice, the next step is to give students voice in the presentation of course material or instruction itself.

Student Evaluations of Teacher Performance

One method for accessing student voice in the educational process is to allow students the opportunity to give feedback on the performance of their teachers which includes giving their

perspective of instruction, organization, classroom environment, and quantity and quality of the amount learned.

Student evaluations in a university setting. Student evaluations have become almost institutionalized at the university setting. Virtually every institution of higher education in the United States implements some type of student evaluation as a method of assessing teacher performance in courses (Onwuegbuzie, Daniel, & Collins, 2009). Student evaluation results are a component of the way colleges and universities determine the quality of an instructor's teaching for purposes of promotion and tenure, as well as for allotting merit pay or raises (McPherson, Jewell, & Kim, 2009).

However, it appears that that student evaluations at the university level originally came about because of the perceived need to increase the quality of teaching. In the early 1990's, research on student evaluations argued that faculty and administrators needed to expand their concept of scholarly work to include not only research and discovery, but also the art and science of teaching (Lattuca & Domagal-Goldman, 2007). Despite the widespread administration of student ratings in the university setting, this original purpose has been overshadowed by controversy. The literature suggests that student evaluations are still a debated topic in the world of higher education, and it would be a rare campus that embraces or agrees on their ability to measure instructional quality (Seldin, 1999; Sanford, 2013).

Arguments against student evaluations. Some faculty members are distrustful of administering student evaluations, perceiving students not to be qualified to evaluate instruction and express concern that ratings are more popularity contests than meaningful feedback. Studies have also questioned differences in how students rate instructors based on variables such as age, sex and expected grades (Spencer & Schmelkin, 2002). McPherson's (2006) research suggests

that higher expected grades had a correlation to higher student rankings and that lower class sized correlated with higher rankings suggesting that increased student learning maybe due to the class size not the instructor. In Zabaleta's study (2007), student ratings are questioned as reliable indicators of student learning and viewed more as indications of customer satisfaction. "Grading leniently, watering down course requirements, refusing to teach required courses, or being an entertainer" (p. 67) are concerns shown in studies of the outcomes of student ratings. "Research has shown that ratings in elective courses are higher than in required courses and ratings in higher-level courses tend to be higher than in lower-level courses" (Ory, 2001, p. 3).

In a famous or infamous Dr. Fox study (Naftulin, Ware, & Donnelly, 1973) still widely referred to, a lively actor, hired to give a lecture in medical school, was highly rated on teaching quality even though the lecture intentionally lacked content. Trout (1997) found that student comments on evaluations often suggested making class more fun. The student need for entertainment factor seems to be a major concern of faculty. Thus, a concern exists that personality may be ranked higher than content or quality of instruction.

Arguments for student evaluations. Despite these findings detailing the concerns over the reliability and validity of student ratings, other studies offer a direct contradiction. Aleamoni (1999) in reviewing over 155 research studies from 1924 to 1998 refuted what he identified as myths perpetuated by faculty to discredit the value of student ratings, indicating that most of the research refuting the reliability of student ratings was written by professors. His research documents studies that contradict the common myths that perpetuate student ratings as inconsistent, popularity contests, lacking mature insights, showing correlation to expected grades or gender, and failing to provide meaningful feedback. His findings show that while warm, friendly or entertaining teachers may be praised by students, these same teachers would be

highly criticized if disorganized or lacking rigor in instruction and that teachers who addressed mid-term student ratings concerns could not only improve end of term ratings but could also improve learning by the end of the term. Likewise, one extensive study on this "excitement" or "like-ability" factor documented the correlation between teacher expressiveness and higher ratings and came to an interesting conclusion. "Investigations of instructional characteristics will lead to more precise descriptions of effective teaching behaviors. These descriptions may help instructors modify their teaching" (Abrami, Leventhal, & Perry, 1982, p. 461). The videotapes used to determine the more entertaining teachers, the ones with the higher ratings, this study now suggested could be used as tools to help less-entertaining professors improve their own teaching or presentation skills. This conclusion reinforces Aleamoni's (1999) summation that student ratings can be used to "enrich and improve the course" as well as to "document instructional effectiveness" (p. 160).

Numerous studies have focused on multiple factors that indeed have the potential to influence course evaluations: class size, instructor rank, instructor sex, student sex, student class year, required vs. elective courses, standard vs. pass/fail grading, major vs. minor courses, and expected grade (Farreras, 2012). However, despite these issues, the evaluations still may hold useful information. Instructors at the University of Vermont purposefully set about using student evaluations throughout a course to revamp and update course content. They concluded that "Detailed student feedback can provide a rich source of information to help instructors evaluate specific elements of course design and structure, make revisions, and assess the effects of those changes" (Brew, 2008, p. 104).

Even if universities chose not to post student evaluations, with the advent of social media, students often take rating their teachers into their own hands. Students have taken student ratings

out of the hands of the faculty, and even the administration entirely with the introduction of new websites, such as ratemyprofessor.com, which allows students to comment publicly and openly on the Internet about their teachers. The website ratemyteachers.com also publishes unsolicited student ratings of public school teachers. Whether at the university or public school level, students, in our communication driven society, may find their own ways to be heard.

Student evaluations in public schools. While the previously cited studies focus on the more common use of student ratings in higher education, there have been studies specifically on the use of student ratings in the public school setting starting as far back as 1896 when in Sioux City, Iowa, grades 2-8, student views concerning what constituted the most helpful teacher and what the characteristics constituted the best teacher were studied (Kratz, 1896). Knight (1922) found high reliability in high school student rankings of general teacher abilities. Light (1930) did a study of high school students asked to rank their teachers from highest to poorest. Of 148 students, 103 ranked the same teacher first. Flinn (1932) found higher inner reliability between student evaluations and administrators.

An early advocate for democratic schools, Remmers (1934) of the Purdue psychology department concluded that high school students not only provide reliable estimates for single traits, but also that they were able to discriminate between teacher traits, and thus overcome the halo effect of being swayed by teacher appearance or personality. Medley and Mitzel (1959), Blanchard (1968), Davidoff (1970) continued to find high reliability in high school student rankings of teacher effectiveness. Pupil Observation Survey (POS) with Student Evaluation of Teaching (SET) (Peck, Olsson, & Green, 1978), Student Observation of Teachers and Teaching Techniques (StOTT), developed by 11th and 12th grade students (Masters, 1977), The Steps to Excellence Questionnaires (Ebmeier, Jenkins, & Crawford, 1991), were all student evaluation

instruments that demonstrated high levels of consistency and reliability. Bryan (1966) using the Pupil Observation Survey had seventh graders concurrently and twelfth graders retrospectively rate the same 34 teachers on nine items using a five-point rating scale. Only four of the 34 teachers received significantly differing ratings from the 12th graders than from the seventh graders, showing then substantial evidence of stability in the student evaluations. In over 20 studies (over almost a 100-year period of time, 1896–1991) evidence found that, in general, personal and demographic characteristics of pupils and teachers have little influence on student ratings and indicated that secondary level student raters can evaluate teachers reliably (Fullman, 1992).

Despite the findings of these studies, implementation and acceptance of student evaluations in public schools is still limited (Kane, 2013). Kauchak, Peterson, and Driscoll (1985) conducted a study on teacher attitude towards student evaluations. One third of teachers said student evaluations were valuable but that professional judgment was necessary in interpreting the evaluations. One third of teachers were cautiously accepting of student evaluations. The last third openly doubted students' ability to provide reliable or valid information about teaching effectiveness. Interestingly, the study also showed that teachers placed little value on administrative evaluations and an even lower value on test scores as being accurate evaluation tools for teacher performance. Schwab and Iwanicki (1988) found that only 11 percent of teachers and only eight percent of administrators strongly supported the use of student *evaluation*. However, 47 percent of the teachers and 50 percent of the administrators did support, with some reservations, the use of student *feedback*. Perhaps, because this study also linked the use of student evaluations to merit pay, support was limited on the use of evaluations for judgment, but more favorable if student input was limited to feedback.

Thus, while secondary students admittedly would be vulnerable to bias, it was concluded in 1992 "that the relevant secondary student rating research does justify the use of secondary students' ratings as one of several criteria in a professional teacher evaluation program" (Fullman, 1992, p. 176). Yet, there was little implementation at the public school level. Recently, however, the extensive research funded by the Bill and Melinda Gates Foundation, the MET Project (2010) has reignited the possibilities of using student evaluations as part of assessing teacher effectiveness. In fact, their research makes the assertion that student feedback is a far better predictor of the quality of a teacher's performance than more traditional indicators of success like principal evaluations and whether or not the teacher had a master's degree.

Arguments against student evaluations. The release of the MET Project findings and its recommendation for the reintroduction of student voice has once again been met with controversy. Author and university professor Stanley Fish responded to the MET Project in this way in an article written for the New York Times (2010, June 21). Taking the voice of a student, he said:

> It may be years before I know whether I got my money's worth, and that goes both ways. A course I absolutely loved may turn out be worthless because the instructor substituted wit and showmanship for an explanation of basic concepts. And a course that left me feeling confused and convinced I had learned very little might turn out to have planted seeds that later grew into mighty trees of understanding.

Public opinion posts on the blog of Fish's article represented some support for his view from teachers. Their input served as inspiration for Fish's second article in the NY Times (2010, June 28). Quoting some teachers in the second article, Fish's article included these sentiments:

> Sorry kids, you are not the authority in the classroom. Me Teacher. You student. Me Teach; You Learn. End of discussion. Education is not a business. You are not my customer. My classroom is not Burger King. You do not get to have it your way. Courtesy and respect does not extend to their ideas, which may or may not be given a hearing depending on the instructor's preferred teaching style, and which may be summarily dismissed if they are judged to be beside the pedagogical point. Treat them as human beings with inherent dignity by all means; but don't treat them as sages before the fact.

No doubt these comments represent the extreme, but the fact that a debate is likely in the public school setting over the reliability of student evaluations is also evident.

Arguments for implementation of student evaluations. However, recent interest in multiple perspective teacher evaluations and the concern over of the lagging performance of American students has generated new interest in student evaluations. With local school districts, 13,600 of them (Toch, 2012, p. 67), seeking for school reform to regain international competiveness, it makes sense that there must be some national direction. As part of that reform, currently, 27 states and Washington, D.C. (see Figure 4) have current student feedback initiatives in progress (Stefanakos, Kesselheim, & Kostin, 2013). While some may consider student evaluations as a punitive evaluation tool, others see student evaluations as a possibility for partnering students and teachers together in a collaborative learning environment.

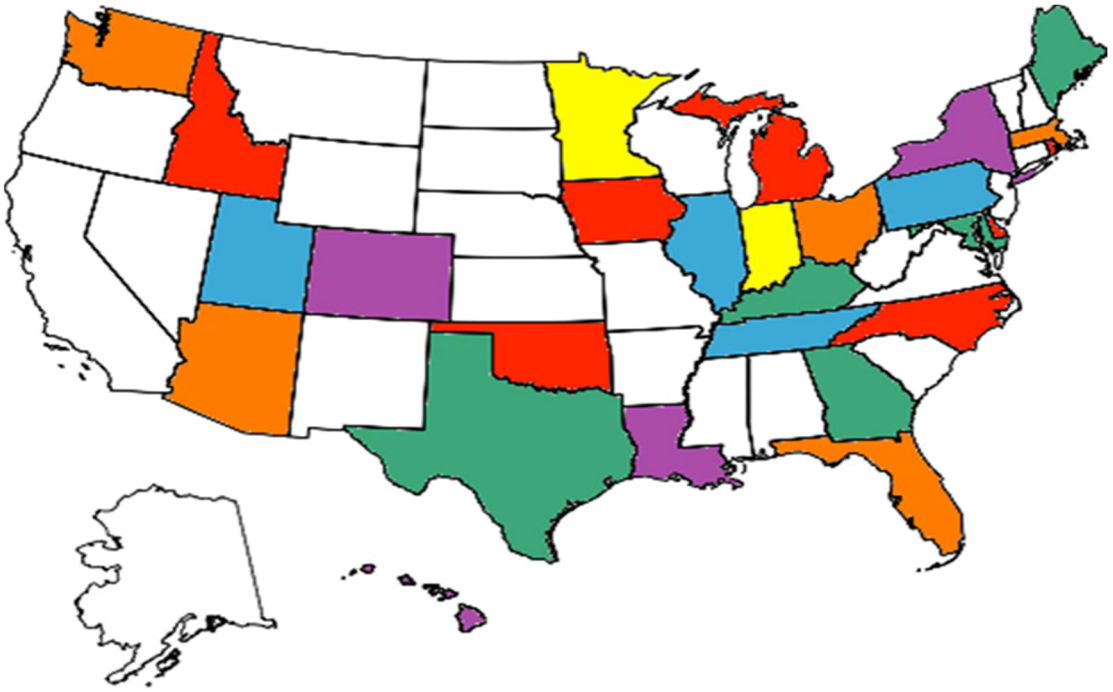


Figure 4. Current states with student feedback initiatives as part of teacher evaluations progress.

MET Project research. The MET Project (2010) has made a case for not only revisiting but revamping teacher evaluations in public schools and seeks to question previous concerns over the lack of reliability of student evaluations as an important piece of the teacher evaluation process. The MET Project comes to this conclusion:

> Reinventing the way we develop and evaluate teachers will require a thorough culture change in our schools. No longer should teachers expect to close the door to their classrooms and go it alone. The quality of instruction is a public good, and improvement will require a collective commitment to excellence in every classroom (p. 8).

The MET (2010) project asserts that collective also means the students. As part of that evaluation process the MET research used observations of over 13,000 lessons taught by teachers and "also asked students to report their perceptions of each teacher's classroom" (p. 2). From these findings comes a simple statement, and proposed validation for the inclusion of student voice.

"The average student knows effective teaching when he or she experiences it" (p. 5). The research found that student perceptions of teacher practices were "remarkably consistent" and that "when students report positive classroom experiences, those classrooms tend to achieve greater learning gains" (p. 5). Student input in the MET Project findings did not demonstrate a popularity contest but rather saw "students give the most credence to a teacher's ability to control the classroom and challenge students with rigor" (p. 5).

Commencing at the same time as the MET Project research, in October 2011, the U.S. Department of Education organized a series listening tours, during which staff members from Obama's administration, in partnership with organizations such as Youth on Board, the Young People's Project, established focus groups to talk about students' need to complete high school and move onto a path of higher education. Students were given the opportunity to vote on the most pressing issues facing education today. Including students in teacher evaluation came in second with over 94% of students agreeing with the need for including student feedback in teacher evaluation (Brenchley, 2011).

Boston Public Schools model. In Boston, research has been conducted on the inclusion of student voice through teacher and school evaluations in successful pilot high school programs with results confirming the MET Project findings of students not fleeing from but seeking challenging and engaging curriculum. These pilot high schools involved student voice and student choice in the decision making process and out-performed other public schools in student achievement (Doyle & Feldman, 2006).

These urban school students have high expectations of themselves and their schools, in addition to clear and articulated needs that they want to see their schools address.

> Through these students' voices, educators and policy makers can hear how they could fill in this missing information and become more effective in their work (p. 394).

The Boston Student Advisory Council implemented a successful campaign to include student feedback in teacher evaluations and practical implementation methods. In an article written as a body of students, the students stated:

> As students, we are the ones in the classroom, and our futures are affected by what happens there every day. Despite this fact, we are rarely provided with the opportunity to give our teachers feedback on classroom management and instruction (Boston Student Advisory Council, 2012, p. 153).

Convincingly, they phrased their argument:

> As young people, we have learned from many of our teachers that they appreciate honest feedback from us so they can use it to improve instruction. And just as teachers seek quality in their teaching, students seek quality in their learning. We want teachers who are creative, engaging, inspirational, and challenging (p. 153).

The recent successful implementation of student evaluations, inspired by the MET project, as part of teacher evaluations in Boston Public Schools in 29 public high schools demonstrated and confirmed the potential offered in the MET Project. The Constructive Feedback (CF) student evaluation process received an overwhelmingly positive response from both teachers and administrators. Teachers who filled out a follow-up questionnaire reported having a better understanding of how their students were learning and wrote that they were working toward adjusting certain teaching practices to better meet the needs of their students (Boston Student Advisory Council, 2012, p. 162).

As a result of a year-long research process and public debate, the Massachusetts Board of Elementary and Secondary Education voted on June 28, 2011, in favor of a teacher evaluation framework that includes student feedback beginning in the 2013-2014 school year. Every student in the state of Massachusetts K-1 2 will contribute to their teachers' performance evaluations. The Massachusetts Department of Education began immediately studying the implementation process with voluntary and pilot program input to develop the student evaluation tool (Boston Student Advisory Council, 2012). Ross Wilson, assistant superintendent over Teacher and Leader Effectiveness of Boston Public Schools noted this about the district's progress:

> All educators in BPS will have k-12 student feedback as a component beginning next year. The feedback will be gathered through electronic surveys. The questions are the same in all schools with the same grade levels. The results are used with other measures to provide an overall rating for every educator (personal communication, February 22, 2013).

Tamika Eastwick, senior project manager for teacher development of Boston Public Schools (personal communication, February 22, 2013), added that the district partnered with Cambridge Education (2013) and used the Tripod Student Engagement Survey offered in hard copy or as an on-line tool. The product is described as:

> The Tripod Project administers surveys with students, teachers and parents to document attitudes, perceptions, experiences, and practices. The resulting data are returned to districts and individual schools in forms suitable to inform and influence deliberations about ways to improve schools, raise achievement, and narrow gaps. Student surveys on classroom conditions, tailored for grade levels K–2, 3–5, and 6–12.

Ron Ferguson who participated in the MET project helped design the surveys for the Tripod Project.

Additional models. Further fueled by or contributing to the MET Project, the conversation around student involvement in teacher evaluations is beginning to expand beyond Massachusetts. Youth in Providence, Rhode Island, are beginning to establish a similar campaign. Alaska implemented a process to elicit formal feedback from both students and parents (Alaska State Statue, 2010), while in the San Francisco Unified School District, middle school and high school students provide feedback on teacher performance and classroom environment (San Francisco Youth Commission, 2009). Schools in Memphis, Tennessee already count student surveys for as much as five percent of a teacher's overall evaluation. Another 11 states, along with multiple individual districts, are in the process of deciding the best way to include student surveys in the evaluations of teachers (Boser & Rosenthal, 2012).

The process of using student evaluations on a new multiple perspective teacher growth indicators has already begun in Los Angeles Unified School District. After a collaborative process, and an initial trial phase, this year is the practice phase, with next year cited for full implementation. On the district's Talent Management Division website (Los Angeles Unified School District, 2012) the following information is provided:

> One of the multiple measures that is being tested out in the EGDC is Stakeholder Feedback, which includes surveys of parents/guardians and students regarding classroom experiences and perceptions at their schools. In December 2012, a series of surveys will be administered that will serve as a source of evidence for teachers and school leaders on the Stakeholder Feedback measure. The survey will be administered to all students in the

EGDC Volunteer Teachers' classrooms (grades 4-12) and offered to parents of the elementary students. Survey results are no stakes and all responses are confidential.

The Los Angeles Unified model stresses the purpose of the new Educator Growth model is to celebrate success and offer targeted support of teachers. Interestingly, the fact that the student surveys are given with "no stakes" or punitive measures during this practice implementation stage is specifically noted. However, the website noted, in the full implementation stage, teacher salaries will be in a *not yet determined* fashion tied to educator performance rankings from all perspectives. Figure 5 identifies the proposed new system for evaluating teacher performance including as one of the multiple perspectives the use of student evaluations.

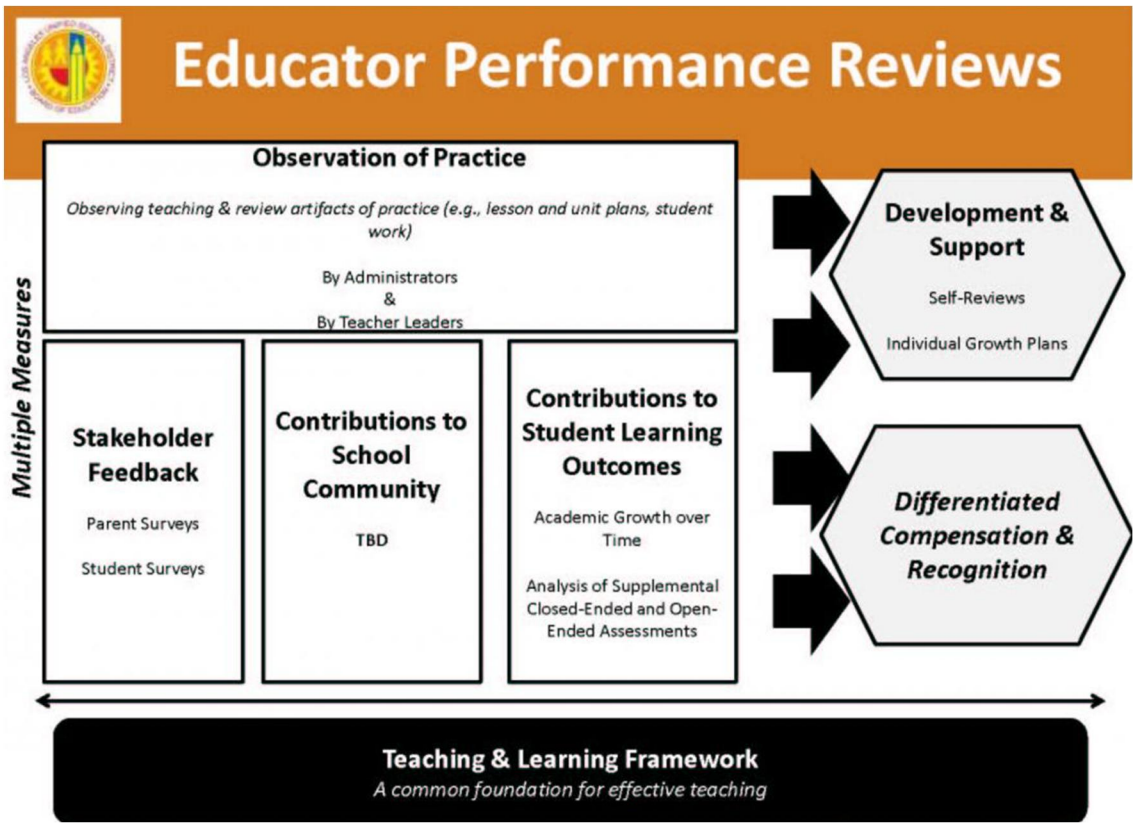


Figure 5. Los Angeles Unified School District multiple perspective Educator Performance Plan. Student surveys are included (Los Angeles Unified School District, 2012).

Methods for student evaluations in public schools. With the possibilities of student evaluations being heralded, research also suggests that student evaluations are just a piece of recommended teacher evaluation procedures and that the process is not perfect, the recommendation was made to use "confidential surveys to collect student feedback on *specific aspects* of a teacher's practice, including those in non-tested grades and subjects" (MET Project, 2010, p. 8). In addressing participants of Voices in Action, Secretary of Education Arne Duncan (2011) said:

> I am a big believer in 360 degree evaluation of teachers… Including students in the evaluation process is one step closer to this 360-degree evaluation, a model in which there are multiple evaluators. In the case of teacher evaluation, these evaluators can be headmasters, other administrators, teachers, students, and parents. The more perspectives there are in the evaluation process, the more comprehensive the evaluations will be, with a higher likelihood of teachers benefiting from them.

Little, Goe, and Bell (2009) created a chart detailing multiple perspectives (see Figure 6) and specific insights each could add to teacher evaluations. The chart includes student evaluations as valuable for specific feedback on non-tested subjects, determining whether a new teacher is meeting classroom expectations, determining the types of assistance a struggling teacher may need, and as the only perspective that gives information relating to how students perceive a teacher's instruction (Little et al., 2009).

Purpose of Evaluation of Teacher Effectiveness	Value-Added	Classroom Observation	Analysis of Artifacts	Portfolios	Teacher Self-Reports	Student Ratings	Other Reports
Find out whether grade-level or instructional teams are meeting specific achievement goals.	x						
Determine whether a teacher's students are meeting achievement growth expectations.	x		x				
Gather information in order to provide new teachers with guidance related to identified strengths and shortcomings.		x	x	x			x
Examine the effectiveness of teachers in lower elementary grades for which no test scores from previous years are available to predict student achievement (required for value-added models).		x	x	x			x
Examine the effectiveness of teachers in nonacademic subjects (e.g., art, music, and physical education).		x		x		x	x
Determine whether a new teacher is meeting performance expectations in the classroom.		x	x	x		x	x
Determine the types of assistance and support a struggling teacher may need.		x	x		x	x	
Gather information to determine what professional development opportunities are needed for individual teachers, instructional teams, grade-level teams, etc.	x	x			x		x
Gather evidence for making contract renewal and tenure decisions.	x	x					x
Determine whether a teacher's performance qualifies him or her for additional compensation or incentive pay (rewards).	x	x					
Gather information on a teacher's ability to work collaboratively with colleagues to evaluate needs of and determine appropriate instruction for at-nsk or struggling students.				x	x		x
Establish whether a teacher is effectively communicating with parents/guardians.				x			x
Determine how students and parents perceive a teacher's instructional efforts.						x	
Determine who would qualify to become a mentor, coach, or teacher leader.	x	x	x	x			x

Figure 6. Matching multiple perspectives of teacher evaluations to specific purpose (Little, Goe, & Bell, 2009. p. 16).

Focusing on solutions, valuing the hard work of good teachers, and "asking the students to help" (p. 8), the MET Project seeks for a collective process for school improvement by identifying and increasing teacher practices that improve student satisfaction, but more important or perhaps essential to that satisfaction, learning outcomes. According to the MET Project, satisfying students the Burger King have-it-your-way method may actually increase student achievement.

Further confirmation of implementing student evaluations comes from a recent longitudinal study on teacher-student relationship (TSR) and school climate (Barile et al., 2012) stating that schools that scored higher on the positive teacher student relationship scale had lower teachers' performances, students may actually believe that teachers care about what they think, which teacher–student relationships may foster a sense of belongingness in school and promote a warm school climate. The more positive teacher–student relationships the greater the opportunity for students to have a voice, which, in-turn, may potentially lead to greater academic success.

Research also suggested that teachers can refer to suggestions from students' comments or recommendations from past evaluations and what specific changes were made because of those evaluations as a way to validate and encourage more meaningful student input and feedback. Students develop a *perceived* control as they view the role that other students played which can lead to a greater sense of actual control in their own learning. Specifically, Perry, Hladkyj, Reinhard, Clifton, and Chipperfield (2005) described perceived control over academic outcomes as "a student's beliefs in his/her capacity to influence and predict achievement outcomes" (p. 536). Overall, they found that students with a perceived academic control had higher GPAs. Evidence again that motivation and student learning can increase with a sense of control or say or voice.

Concerns over potential punitive use of student evaluations in public schools. With the increased use of student evaluations at a university level, the controversy over the evaluation process has also increased. "The reason for the controversy is not just the student evaluations themselves, but rather the way the student evaluations are often used" (Adams, 1997, p. 10). Student evaluations often serve as a summative tool used in administrative decisions about faculty tenure, promotion, and merit pay (Kulik, 2001). This is the same debate heating up in the public school setting. Aleamoni (1999) showed that student ratings have disadvantages of potential misuse or misinterpretation by those who use ratings for punitive purposes.

> Too often, SET (Student Evaluation of Teaching) systems have been compulsory, publicly displayed, uncontextualised, unsupported, simplistic, and interpreted in isolated ways, features which render SETs punitive bureaucratic tools rather than supportive mechanisms through which enhanced learning environments can be created and sustained (Moore & Kuol, 2005, p. 148).

Kozub (2010) concludes that administrators should interpret student evaluations with care, as many factors play into student ratings including if a faculty member teaches at an unpopular time of day, in areas that some students do not find intrinsically interesting, or if the instructor has been given the reputation of being a difficult grader.

Formative vs. summative use. Teachers then justifiably fear the use of evaluations as summative rather than formative assessments that are used solely or without a chance for improvement to determine merit pay, tenure, or retention. If reduced to this purpose, student evaluations to improve instruction may erode to embarrassment or humiliation as the vehicle to drive the change of practice rather than using partnership feedback to inform practice. The original purpose of student evaluations of university faculty was not for a summative, personnel

factor but as a formative assessment — that is, for the informing of improvement of individual instructors' teaching (Blunt, 1991) stated that the original intent is still considered important. Onwuegbizie et al. (2009) found that student evaluations have an important assessment function if used as a formative measure by faculty looking to improve their teaching skills and course design.

The MET Project suggested that the same purpose for the use of student evaluations as a source of formative feedback. Daley felt that this distinction in the purpose of student evaluations is crucial to their wise implementation.

> Once we make clear the context of the student opinions we solicit, teacher evaluation instruments will serve teachers as well as students and administrators. Conversely, it is unfair to perpetuate a system that benefits students and administrators while threatening faculty members (Daley, 1999, p. 64).

Low stakes vs. high stake settings. In fact, one of the arguments questioning the conclusions of the MET project findings (Rothstein & Matis, 2013) states that the data was obtained in a relatively low stakes setting. In other words, the teachers participating in the MET project research submitted to the evaluations with no specific consequences for faring poorly. Rothstein and Mathis refer to Campbell's Law, "that a measure that performs well in a low-stakes setting will inevitably be distorted when the stakes are raised," and "many analyses of the design of performance evaluation systems recommend that the stakes be kept relatively low, to reduce the incentive to distort the performance measures" (p. 10). Thus, they question if the MET project results would be the same under a high stakes system. They go on to further state that the MET project "cannot inform us about the sensitivity of teacher evaluations to the distortion efforts that will inevitably arise from higher stakes" (p. 10).

Performance or merit pay concerns. It is true that the US Department of Education, Arnie Duncan, and the federal government's Race to the Top initiative all call for high stakes-performance pay plans (Woessmann, 2011). Therefore, while the stakes may be low now, teachers understandably fear the possibility of high stakes ahead.

After four years of researching elementary and secondary public school merit pay programs, Ritter (2010) summarizes the two camps on this issue:

> Passionate promoters of merit pay have argued that such a scheme would properly align incentives for teachers so the most talented are recruited, the best are rewarded, and the laggards are relocated to a different profession. On the other hand, fervent foes of the practice contend that performance pay would not capture all that teachers do and would instead result in a counter-productive narrowing of the teacher's goals and divisive competition between and among educators who would otherwise seek fruitful collaboration (p. 32).

Ritter comes to this conclusion, "A review of empirical data from the handful of merit pay schemes in the United States and abroad revealed, not surprisingly, that the results of these programs are mixed" (p. 32). Montgomery County Education Association in Maryland found greater success in stakeholder partnering by recently doing away with a merit pay system (Sullivan, 2012). With mixed reviews and few proven systems, merit pay plans based on formula (many with formulas still yet to be undetermined as is the case in Utah) on teacher evaluations with each perspective weighted, including student evaluations, the uncertainty and fears over the process are heightened.

Public display issues. Anxiety also exists about the possible public display of teacher evaluation scores. Again, with complex differences in the type of students and classes taught,

this system assumes an *apples to apples* comparison that does not exist. According to the Sunshine Review (2012), a non-profit organization advocating government transparency, "Faculty evaluations on an individual basis are generally not available through public records . . . many will not disclose results of individual faculty member evaluations" (p. 1). Their article goes on to state that in public schools only nine out of the 50 states so far require public publication (most of these newly instated laws) of teacher evaluations with six allowing for partial or unidentified disclosure, six others use a balancing test, Utah leaves it up to the legislature, and others have no law specifically stating one way or the other. The balancing laws such as the one used in Wisconsin for the University of Wisconsin and public schools (University of Wisconsin Handbook, 2012) state as follows:

> The Attorney General has ruled that student evaluations of instruction, whether viewed as personnel records or not, are records subject to the Wisconsin Public Records Law. This law presumes complete access to all records kept by a state agency, including complete access to personnel records. Denial of public access to any record is presumed to be contrary to the public interest and only in limited circumstances may access be denied. Denial is permitted only where the interest of the public in prohibiting access outweighs the interest of the public in obtaining access (p. 14).

The last line of this law represents the crucial dilemma of publishing student evaluations. Could publication of student evaluations be denied based on evidence that outweighs the need for public disclosure? Would publication threaten to not only undo strides in making teaching more collaborative, but also threaten to comprise the process or purpose of student evaluations in the first place, placing the emphasis on punitive punishment rather than collaborative support? Is it counter-productive to enter into a collaborative student/teacher partnership and then have

student voice used for punishment rather than partnership? These are some of the issues to consider in increasing teacher anxiety over the publication of student evaluations.

Evaluation instrument design problems. Another major concern over the implementation of student evaluations is the questions themselves. Dmrtiry (2012) suggested that the student evaluations should utilize a standard scale, such as a five-point Likert scale in order to facilitate comparison of each question and that the scale should be comparable across the questions on the form.

With validity issues related to the level of student understanding and therefore ability to judge pedagogy and curriculum decisions, the actual content of the questions the students are asked in a teacher evaluation are at issue. The very wording of questions may also prompt student bias. However, beginning in the 1970's and continuing to the MET Project, several student evaluation instruments were developed with an emphasis on non-biased language and a focus only on elements that students are in an overall position to judge. Marsh and Hocevar (1991) suggested that Frey's Endeavor Instrument, The Student Description of Teaching (SDT), Marsh's Student Evaluation of Educational Quality (SEEQ), Michigan State SIRS Instrument were reliable student evaluation tools. The similar qualities of each of these instruments call for student judgment limited to teacher enthusiasm, organization, instructor's interest in student learning, group interaction, appeal to student interest, grading procedures, and course structure (breadth and workload) (Marsh & Hocevar, 1991). The authors felt these characteristics were within the realm of student capabilities to determine from uniquely their perspective as the receivers of the teacher instruction.

In referring to the work of Marsh of evaluating student evaluations of quality teacher instruction of 195 teachers over a thirteen year period, Richardson (2005) made this renewed

conclusion: The inter-rater reliability of the average ratings given by groups of students is high, provided that the average is based on 10 or more students. Evaluations of the same teachers given by successive cohorts of students are highly stable over time (Richardson, 2005).

The MET Project (2010) also developed a student evaluation tool again specifically designed to ask student accessible questions about their classroom experience. The survey identified seven constructs, the seven Cs, that are "core to a student's experience in his or her classroom" (p. 5). The seven Cs, care, control, clarify, challenge, captivate, confer, and consolidate and bear similarities to the kind of traits measured in earlier instruments. The survey data showed that students were able to differentiate between teachers and their classroom environments. Even when a teacher taught multiple classes, the perception of students from different classes was remarkably consistent.

Focusing on the questions used in the student evaluation can increase the reliability and usefulness of the information.

> When students are asked questions that probe learning (such as whether their attitudes or beliefs were changed, whether they understand connections more fully, whether they feel more confident in their ability to tackle problems of the field) and the teacher's approach to facilitating learning (such as whether the teacher answered students' questions, invited students to office hours, and promoted student engagement). The evaluation comes much closer to measuring teaching skill (Calkins & Micari, 2010, p. 18).

Thus, the survey instrument itself may be suspect without necessarily the process of collecting student evaluations.

The MET Project does concede that an instrument tool can be distorted if students do not take the survey seriously or do not believe that their answers will be kept confidential. Thus, the

project calls for multiple methods of feedback for teachers and checking one against the other to determine the accuracy of not only the measurement tools but the implementation of the tools.

Given the heightened interest in the use of student surveys across the nation, many organizations are currently putting money and research into trying to improve the quality, usefulness, and affordability of student surveys, including Tripod Education Partners, Panorama Education (financial backing from Facebook), Cambridge Education, My Student Survey, Battelle for Kids/Gallup (Cavanaugh, 2014).

Recently, Onwuegbuzie et al. (2009) discussed a meta-validity model recommending that STE (Student Teacher Evaluations) results be used as only one indicator of teaching effectiveness, and Chulkov and Van Alstine (2012) adds that other lines of evidence could include portfolios, classroom observations, peer observations, student work samples, teacher self -reflection, value added test scores, and course design. "The potential for bias, however, in no way renders the evaluations useless. The answer is not to do away with them, but rather to use them wisely" (Calkins & Micari, 2010, p. 18).

Creating a safe environment. For 16 years, Chris Unger (2003) worked at the Harvard Graduate School of Education in *Project Zero* where he developed new programs of learning and studied whether the learning programs worked, or did not work. He started to interview students about their perspectives on their classes and teachers and was surprised about how much they had to say. Although not part of his research, he felt compelled to let teachers view these interviews. He found that usually the teachers who were already doing well but who were also passionate about continuing to improve were the ones who wanted to see the video- taped interviews. But he also noted that much of the receptivity depended on whether or not the school had created a safe and positive environment for teachers as well as for students. His suggestion

for creating a teacher-student partnership in the learning process was to allow for a safe environment to facilitate collaborative sharing. If teachers do not feel threatened by hearing the feedback of their students, then the way is open for collaborative efforts to make positive changes not only in classrooms, but in the whole school environment.

Implementation ideas for student evaluation. In Bursch (2007), he suggests that public education has become institutionalized in practice, slow to change because of the strength of cultural norms. He poses a crucial question: "Why is it that, in education, there are many examples of reforms that have been faithfully implemented and yet very few examples of sustained improvements at the core of schooling?" (p. 84).

Using force field analysis. Implementing change in education can be a difficult. Kurt Lewin (1951) developed force field analysis as a method for implementing needed organizational change. By evaluating the forces driving change as well as the forces resisting those changes, a plan for addressing the resisting forces is necessary for change to occur. Before change can be facilitated, one force has to be strengthened or one weakened for the hold to be broken or unfrozen (Owen, 1998, p. 315). Owen suggested that his type of analysis can be applied to educational change.

The force-field analysis is diagnostic: "it permits the preparation of plans for specific action designed to achieve changes sought" (Owen, 1998, p. 314). The following diagram could represent the use of the Force Field Analysis applied to the implementation of student evaluations (see Figure 7). The implementation of student evaluations could be approached by examining the driving forces and resisting forces.

Powerful resisting forces are caused in this situation by distrust of the process and legislative mandates, concerns over job stability, merit pay, anxiety over student judgments

being based on personality factors, and fear of public humiliation through public publication of student feedback while powerful driving forces push from the other side as in new research, teacher value of feedback, and increasing student motivation and voice through facilitating a partnering in their own education. Successfully incorporating student evaluations into the "norm" of teacher and school practice may be achieved by addressing the driving forces while finding ways to weaken or alleviate the resisting forces or in this case the concerns that act as the resistors to student evaluations.

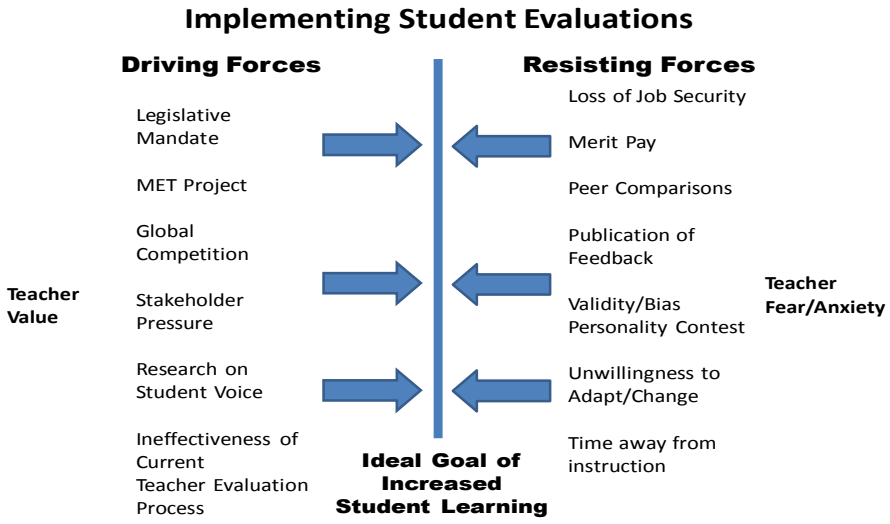


Figure 7. Force field analysis of implementing student evaluations.

The "unfreezing" catalyst may be legislative mandates. Yet, that legislative pressure could be seen as an opportunity to remold in an improved form. Unless teachers and administrators along with legislatures have a common higher purpose, it may be like trying to get an atheist and a true believer to run a church together. For success in that reshaping, every role should be defined and related to the mission of increasing student learning. All stakeholders should be working with the same goal in mind (Moldof, 1993): increased student learning for all.

Some teachers are in a status quo. They have never had to ask their students, their clients, for their opinion, but by listening, new opportunities and yes, new threats are both possible. An improved teacher/student relationship could also increase job satisfaction and enjoyment for teachers. Owens (1998) stated that with "Unfreezing" or breaking the equilibrium, change can be "traumatic" (p. 313) at first, but once it is achieve, it can be part of a natural process (see Figure 8).

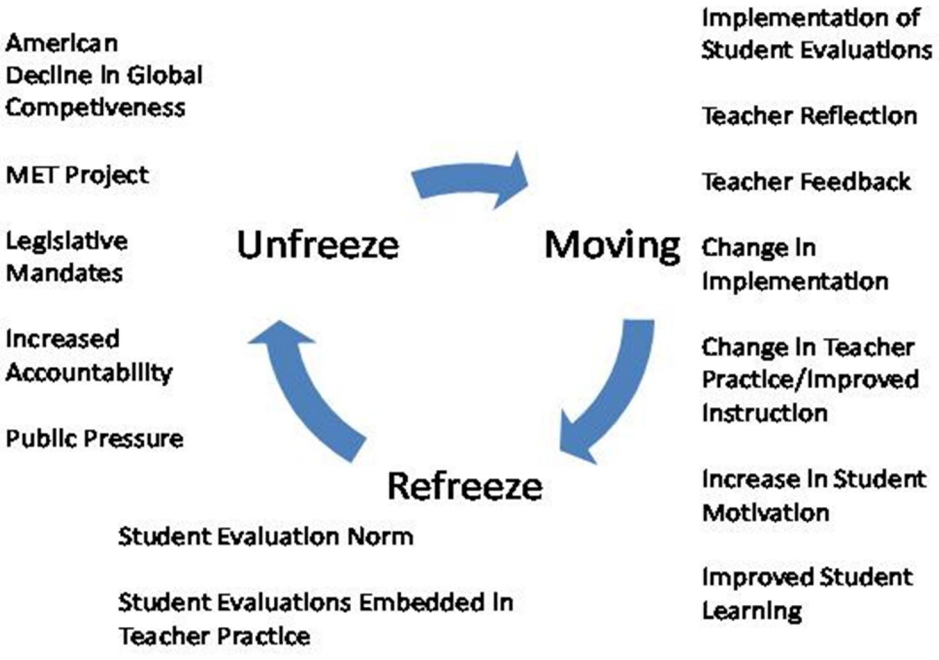


Figure 8. Unfreezing teacher practice (Owen, 1998).

Owens (1998) also warned that if change is forced by additional administrative pressure along with, in this case, legislative pressure, the counter pressure of the resistance may also increase. Using an open dialogue, allowing teachers "the right to question" (p. 315) and helping them to deal with the change with a non-judgmental, problem-solving approach that values maintaining personal dignity is the best way to diminish teacher resistance.

Collaborative involvement in student evaluation process. Thus, many of the roadblocks to teachers finding student evaluations helpful need to be overcome. Teachers, too, need to have a say in how student evaluations will be administered or their motivation to make the process meaningful could by compromised. Otherwise, the value of the student feedback may be inhibited.

Teacher ownership facilitated by being part of a collaborative implementation process may ease anxiety. In the case of Utah legislation, administrators will also face new accountability evaluations. Public Education Employment Reform (2012), Utah House Bill 64 requires that an administrator's evaluation include components such as student achievement, parent, student and employee input, and the effectiveness of evaluating employee performance (lines 860-867) and implements a compensation system for administrator's as well partially based on an administrator's evaluation (lines 875-880). So there can be a sense of *we are all in this together*.

With individual school needs in mind, Marsh and Hocevar (1991) also recommended that student surveys could be self-developed by schools beginning with a review of research-based survey instruments and then including student, teacher, administrator, parent, and district feedback into the development of a student evaluation instrument. A process designed through a collaborative and democratic process that includes feedback for potential usefulness deemed appropriate by the specific stakeholders involved could be a means to address concern over validity and by creating ownership for the instrument itself may alleviate some teacher anxiety. The new evaluation process could be supported as learning by doing approach while new behaviors are given the chance to develop and old ones slowly melt (Owen, 1998) rather than forcing teachers and administrators into a massive melt-down.

Training teachers and students on the process. Administrators could facilitate training and focus group discussions that help teachers and administrators better interpret student feedback. The purpose of student feedback is to gain insights into the student perspective. Although that student perspective may difficult for teachers to take, knowing what the students perspective is can be the very way of impacting, changing or addressing, that perspective. The teacher's negative attitude toward seeking student input may be part of students' concerns or they very catalyst for negative student feedback. "Teaching is a very personal act, and it is hard to accept criticism of something so close to our essence. But if we cannot or if we react defensively, we destroy all hope of getting honest and useful student feedback from that class again" (Svinicki, 2001, p. 17).

On their Center for Learning webpage, Vanderbilt University posted an article entitled *Making Sense of Student Evaluation Feedback* (2012) that offers training in the form of tips for teachers when considering their student evaluation results. The very posting of this document represents not only the university's attempt to train teachers in receiving feedback, but also the value or importance that the university places on the training of teachers to receive and understand the purpose of feedback as part of the student evaluation process. The advice offered addresses specific ways for teachers to make feedback less personal and more reflective.

- Look for patterns in students' comments—identify trends, note what you have done well and what needs improvement.
- Take your experience into account. If you are new to teaching, the school, or even the course, you may still be learning about various aspects of being a professor, such as course design, teaching skills, student interaction, and departmental expectations.

- Take the context and characteristics of your course into account. Research shows that student evaluations often are more positive in courses that are smaller rather than larger, and elective rather than required. Also, evaluations are usually more positive in courses in which students tend to do well.
- When dealing with negative student feedback: Know that almost all faculty members receive negative feedback at some point in their careers, including those who are senior and highly successful.
- Allow yourself to acknowledge that it can feel hurtful or make you angry, but also provides a pointer toward important areas for your continued development.
- Bear in mind the most frequently mentioned areas for teaching improvement in analysis of student evaluations within and across universities: (a) clearer, more specific in-class communication and (b) clearer, more explicit organization of course content.

Students, too, can be trained in the art of giving feedback and given the "opportunity to see models of good feedback" (Lewis, 2001, p. 18). Rather than saying that the instructor was "so disorganized," students can learn to be more useful by enumerating the observations that led to such a label.

> Communicating expectations about the feedback is often enough to influence the amounts and kinds of information the students think to give. Because the skill of giving feedback is becoming a more and more important one as we move toward teamwork in classes as well as the workplace, one possible solution to the problem of desultory student feedback is to take the time in class to teach students how to give feedback (p. 19).

Including the opportunity for scaled scored questions as well as open-ended comments increases the opportunity for teachers to learn from student responses as well as the opportunity for students to uniquely express their opinions.

Faculties have shown a solid preference for student written comments despite the common confusion and contradictions that the comments sometimes elicit (Svinicki, 2001). Invariably, some students may say, "You taught the most wonderful class I have ever taken" and others (in the same class) may say, "You are such a terrible teacher that you should be fired!" Many instructors may be tempted to dismiss the important information these comments provide about their teaching and their students' learning by using such comments as evidence of the unreliability of student feedback. Within the comments section can be simple revelations that justify not throwing all comments out with the bathwater in order to find the simple but often profound treasures of information they can provide (Lewis, 2001). Chulkov and Van Alstine (2012) summarized that the use of written comments enhances the assessment information available to the instructors and can help improve course design and teaching effectiveness.

Students tend to write more and provide more constructive comments when they are provided with prompts for the open-ended questions rather than just a section labeled comments. Prompts can also have the added advantage of keeping instructors from overreacting to a single negative comment, a constant of human behavior in reaction to evaluation, because there will be multiple responses to the same prompt (Lewis, 2001).

The use of or including student self-reflection questions as part of the student evaluation of teachers on their own learning and studying practices can be one way to help students focus on the partnership of the learning process. This self-reflection would help students take more ownership of their education and also reduce the potential for teacher bashing. In fact, students

of the Boston Student Advisory committee suggested the addition of a student reflection on the student's efforts in the class as a way of demonstrating that their committee understands that the responsibility for learning and teaching has to go both ways.

> Favoring easy teachers and penalizing demanding teachers was a huge concern from many of the people with whom we met. In order to alleviate this concern, we decided it was important to evaluate ourselves, too. If we could not honestly and openly respond to questions about our own learning, then perhaps we could not honestly provide feedback to our instructors (Boston Student Advisory, 2012, p. 155).

Wright suggests that student evaluations, especially in the public school setting, be confidential rather than anonymous (Wright, 2006). Under a system of complete anonymity, students need take no responsibility for their opinions. With no possibility for follow-up, students need not think through their responses. If the evaluation results could be tracked by an administrator, follow-up would be possible for serious student concerns about teachers or serious concerns about a student response. Students might not feel as free to use vulgar or strong language. While protecting the students' right to remain anonymous to the teacher to ensure honest feedback, there would remain some way to find out more information if needed. A tracking method also makes it easier when using an on-line program to make sure students only take the survey one time for each teacher. However, the fear of repercussion may be enough to squelch the very student voice the process is seeking. As a collaborative partnership develops and cooperative learning environment is created, hopefully, the concern over anonymity may decrease.

As a model for that collaborative effort, teachers can always provide opportunities for student voice long before the final student evaluation is administered. If the students are given an

opportunity to provide early feedback and they see that their feedback is acted on in a positive way, this signals that the instructor values student input. It can be effective to also refer to, as previously stated, changes made in instruction due to prior student feedback or explain the reasons or necessity for certain practices remaining unchanged. These measures, if implemented to foster the student-teacher partnership early on in the year and not to purposefully alter final evaluations in the end, a process that may backfire, teachers can potentially decrease negative final student evaluations and increase student learning and motivation early on (Svinicki, 2001).

Student feedback can be obtained in many ways other than through the administration of formal questionnaires and indeed many teachers elicit student feedback routinely about tests, units, projects, etc. throughout a course. Good practice and common sense would still encourage the use of all these means of encouraging student participation throughout the year to maintain and enhance the quality of teaching and learning in education.

Benefits of formal evaluation process. Why then administer formal student evaluations? Richardson (2005) concludes that formal administration of student evaluations provides an "opportunity to obtain feedback from the entire population of students; and then document the experiences of the student population in a more or less systematic way" (p. 401). There would seem to be merit in a uniform method or school-wide system for obtaining feedback on teacher performance for reliability and validity purposes just as administering standardized tests or common assessments have their place for assessing students.

Summary

The revisiting of student voice, therefore, takes on a different focus than the movement for student rights of the 60's and 70's. Today, democratic leadership in schools is being defined as distributed leadership which accesses multiple perspectives and multiple stakeholders. With

the threat of American students losing competitive ground in a global marketplace, stakeholders are calling for increased accountability measures of public schools including revisions in the ratings process of teacher performance. Many federal and state mandates now require a teacher evaluation process that includes multiple perspectives even adding the voice of the students.

With education being given the responsibility for the future global economic success of its students and the kind of higher level of creative thinking required for that achievement, it makes little sense to prepare students for the future by putting off their actual and meaningful participation in the present. In the words of the Boston Student Advisory Council:

> Being the primary consumers of education and spending so much time in the classroom make students prime candidates to be evaluators of their learning experiences. Our education determines our future; let us have a say in our futures. . . Students and teachers should share the responsibility for our education by working as equal contributors in the classroom (p. 154).

Student evaluations offer one method for obtaining student voice and a uniquely student perspective on improving teacher instruction. Claims about the reliability of SETs might not ever be possible or desirable as reliability is generally considered a product of the data and not the instrument" (Smith, 2008) and the context of the data differs in every situation. While reliability and validity issues cannot be eliminated and biased and hurtful student ratings for a variety of reasons will continue to exist even for good teachers, this may not be enough to discredit the voice of the rest of the students nor to keep teachers from student feedback on their instruction. However, teachers, too, have legitimate fears over the way that student ratings will be embraced for personnel decisions, merit pay, published, or used as a summative single judgment rather than just one source of formative feedback on their performance. This research seeks to find

possibilities in the process of implementing student evaluations of teacher performance that may minimize teacher concerns as well as maximize the method as a way to seat students together with their teachers around the stakeholder table.

Chapter 3

Methods

This study was designed to view the value of the student evaluation process from the perspective of the teachers rather than looking at the validity or reliability of the actual student input. This chapter outlines the specific use of both qualitative and quantitative analyses, a mixed-method methodology, to analyze the teachers' perspective through the access of archival case study data containing teacher questionnaires and focus group discussions on the first time implementation of two phases with the addition of a non-archival third phase over the period of three years of student evaluations in a specific high school in Utah**.**

Phase I Archival Case Study

The proposal. In initiating organizational change, leaders can stabilize employee morale by instigating a sense of "belongingness and creating participation in decision making" (Lunenberg & Ornstein, 2004, p. 9). As the assistant principal of the high school in the archival study, I (the researcher) sensed the timing for implementing student evaluations as a collaborative process was an immediate opportunity. If our administration waited until HB 64 mandated change, the potential of creating a collaborative partnership for initiating the process would be lost, in other words process by choice not mandate can be empowering rather than power driven. Working with student leadership was my area of expertise and responsibility, making the imminent requirement for seeking student voices a particular interest. However, the cause would not overshadow the teachers' legitimate concerns which I also shared.

"Enthusiasm is not the enemy of wisdom" (Weick, 1987, p. 231). It can be action that provides the material for planning. Instead of waiting for experience to teach, Weick says that action can create experience. Using what he calls the "just-in-time" strategy (p. 229), Weick

suggests that having a large repertoire of skills and resources from which to act quickly even improvise with when necessary is strength. There was no time for long-term strategic planning. House Bill 64 was in the process of passing in the Utah House. The busy end of the school year was fast approaching. If there was to be a trial run and then time for that initial test run to inform a second trial the following year, student evaluations would have to be given within six weeks, after end of year testing and before the last week of school.

Instead of a potential for chaos, improvisation is seen as a way to produce structure. Hesitancy in the face of disorder can actually perpetuate further chaos. "When people act, they absorb uncertainty" (Weick, 1987, p. 230). Thus, having the courage to forge ahead rather than to take time to ponder a plan was my strategy. "Optimism may not be the denial of reality. Instead it may be the belief that makes reality possible" (p. 231). Weick added justification to my line of thinking and clarified the use of positive thinking as a method not just as a vehicle for happy thoughts, but as an incentive for action and thus an opportunity to discover or make meaning together as a faculty. "Meaning lies in a path of action. . . If you get people moving, thinking clearly, and watching closely, events become more meaningful" (pp. 222–223).

With the approval of the district and school administration, I prepared a presentation on the potential positive outcomes of student voice and the inevitability of the impending legislation and launched the proposal for a trial run of the student evaluation process at faculty meeting in March 2012.

Study design. The student evaluations were to be completely anonymous, with no identifying features attached that would link them to students. This decision was never really discussed just accepted as the norm, a way to elicit honest feedback without the fear of retaliation or maybe because this is the way that university evaluations are conducted. The

principal and teacher would be able to see the responses, but would not be able to identify the student.

The decision was made to use an on-line survey tool even though the task of getting almost 1200 students through 8 evaluations with a limited number of computer labs, 6 in total with 1 portable lab, but only 2 that were not also used as classrooms, was daunting. Paper and pencil or scantron and pencil would be easier to administer in one shot, but also would not facilitate an open-ended comment section at the end nor the use of immediate access to results including charts and graphs to inform teachers of class wide results at a click. On-line tools do have a cost, but so do scantrons. With the pending time crunch, using Survey Monkey was chosen because of the low cost (there was no budget) and ease in already knowing the program by the principal.

Each teacher was given an individual link that students could access but in standard format to be easily remembered, teacher's last name and first name with school acronym and surveymonkey.com. Results were protected by individual teacher passwords. The principal could go in and open and close the link to hopefully keep students from taking the survey more than once, but this was not guaranteed because there was no way of identifying respondents or of locking respondents out. It was also thought to be a more uniform, standardized method if students took the survey under a similar, supervised atmosphere assuring individual not group responses.

Collaborative creation of evaluation tool. With basics in place, all that was left was the tough part, now that the faculty was on board: the creation of the evaluation tool or questions and eventually an implementation schedule. The process would qualify as backwards design or beginning with the end in mind. True to the collaboration principle, a committee of faculty and

administrative volunteers, open to any who wished to participate with lunch provided as the only incentive was formed to draw up a draft or proposal of evaluation questions. Lunch meetings would allow coaches and other extra-curricular advisors to participate. The committee ended up being seven members, two administrators and five teachers, each teacher was from a different department, four male, three females, three new teachers (less than five years) and four veterans. The upcoming draft would then be given to the faculty as a whole, the School Community Council parents, and students in committee members' classes, and student council for additional stakeholder input. The idea was to create questions with our students and our needs and purposes in mind while still using researched based questions as starting or talking points.

In our initial meeting, the committee decided to follow the model of the MET Project (2010): keep the list of questions short, limited to ten if possible, write the questions in student-friendly language, target questions that would be designed to give useful, specific information to teachers, and to develop questions that students had the capacity to answer. From the beginning, the committee also wanted to include a place for additional written comments at the end, acknowledging that here was the possibility for the most helpful and most hurtful feedback. But there was a willingness to take the risk in order to access the most valuable information, both from the committee and the faculty as a whole.

To the next meeting committee members brought ideas for questions. Some brought hand written-ideas and some lists of university student evaluation questions. Both the principal and I brought researched based questions from professional journals including those used in the MET project. Commonalities among the questions were identified and possible choices discussed.

I prepared a new list of the committee's choices while eliminating overlapping questions, giving the list to committee members ahead of time so that they could come to the next meeting

prepared for a final discussion. Teachers shared the list with their students gathering their input and suggestions to bring to the next meeting as well. I also presented the list and the proposal to the School Community Council parents and the student council to allow their feedback to also be added to the discussion. From these efforts, at the next meeting, the committee reduced the list of questions to fourteen. All agreed that we were not willing to give up any questions at this point, so we moved on to focusing on the specific wording of each question, aiming at simplifying. With a crafted list of questions ready, I presented the committee's product at faculty meeting the end of April 2012.

Escaping the committee's now weary eyes, one faculty member pointed out that most of the questions were worded from a positive stance, while two used a presumed negative stance. For example, the first question stated, "My teacher does not waste class time," which had a negative base. The suggestion was to change the question to "My teacher uses class time well." Since all questions were going to be ranked on an agree/disagree seven point Likert scale suggested by the research for standardizing answers to allow for common comparisons, it would make more sense to have all questions asking for the same ranking approach as to how well teachers did or did not fulfill the positive presumption. Despite the committee's best intention, this crucial critique had been overlooked, but this faulty at large input almost in and of itself reinforced the collective nature of our project.

- My teacher uses class time efficiently (no wasted time).
- When I am confused, my teacher helps me get straightened out.
- I believe that my teacher wants me to succeed.
- My teacher is fair and consistent.
- My teacher enjoys teaching.

- My teacher acts like a professional.
- I know what my teacher wants me to learn each class period.
- I am learning what I should be learning in this class.
- My teacher encourages active student involvement and participation in the learning.
- My teacher keeps me updated on my progress.
- My teacher responds to all students respectfully.
- My teacher is available for me to get help outside of class.
- My teacher demonstrates good knowledge of the subject material of this class.
- My teacher sincerely cares about me as a person.

The 7-point Likert scale was used to represent a range of responses as opposed to just a yes or no response. After the discussion on whether forcing a choice was a better option by using an even scale that does not allow for the "neither" or "undecided" response, the decision was made to use the uneven scale and allow students, especially if they did not feel the question applied or did not understand the question, to remain "undecided" which we hoped would lead to less student frustration. The 7-point scale rather than the 5-point scale offered a greater range in responses and while lengthening student choices it was thought by the committee to offer more specific feedback. The 7-point scale went from one (strongly disagree) to seven (strongly agree).

Using prompts for the open-ended questions, the following two questions represented the final decision as one question called for a positive response and the other could be seen as asking for a more negative response.

- What I liked best about this class was . . .
- If I could change something about this class, it would be . . .

At the end of the questions evaluating their teachers, a student self-evaluation was added asking students to use the same 7-point Likert scale to evaluate their own effort for the class.

- I spend time studying outside of class.
- I completed all homework assignments for this class.
- I rarely miss this class.
- I put forth my best effort in this class.

The faculty approved the list and there was a renewed sense of purpose and ownership which helped with addressing the high level of teacher flexibility that would be needed to accommodate the implementation process in a two and a half week time frame, starting the second week of May. Teachers would have almost no access to the computer labs and had to be willing to be interrupted throughout the day. Exact times would be hard to determine. The faculty agreed to let me know the times or days it would not be convenient for me to come, many still giving final tests, even though school wide end-of level testing was completed. Otherwise, it was accepted that the protocol would be to just be glad to see me when I got there. With more time, a better schedule could be created- even a Google doc sign-up.

In one week, the principal set up the questions in Survey Monkey with an individual survey created for each teacher. Questions could be copied and pasted in, but the process was still time consuming. With a student evaluation tool in place, this process would not need to be repeated every year, only tweaked to add or delete teachers.

Implementation Phase I. To try and standardize the implementation process even further, an administrator, mainly me, would come into each teacher's class and take the students to the computer lab, facilitate the student evaluation with the students for that particular teacher, return them to class and begin the process all over again. Students soon were able to complete

the surveys quickly eventually being familiar with and hopefully not sick of the routine. The amount of time and administrative involvement and availability is definitely a factor that also needs to be considered in the implementation process.

Here, as another a side note, while the actual response of the students is not the focus of this study, it is important to note due to the key component of the potential in increasing student engagement by accessing student voice that is part of this study that even, I, a strong supporter of advocating for student involvement, was shocked at the intensely, overwhelming positive response from the students, *all* students. It is completely accurate to say, that at least to my face, I did not have one student in the entire school refuse to go or even make a negative remark about participating in the process.

Time was running out; it was clear that some classes may be missed, so the goal became to ensure that at least 75% of all teachers' classes had been surveyed. Most teachers had all classes surveyed, but 20% of teachers only reached that 75% goal. Results were immediate, so teachers were able to see the response of their students the day the surveys closed. With just a few days left of school, teachers viewed their feedback. Somber and reflective moods ensued when evaluation became the topic. It was clear that while most probably received overall good reviews, that few had been left unscathed. The faculty was definitely feeling the effects of end-of-year exhaustion now mixed with the sting of a few hurtful comments. However, teachers continued on by filling out the teacher questionnaires after a faculty meeting and participating in focus group discussion during PLC meetings.

Phase II Archival Case Study

Review, revise, revamp. The following school year, I presented a brief summary of the responses in a faculty meeting as well as the following proposed changes to be included in Phase II. This meeting also oriented new faculty member as to the plan.

Using the responses from the open ended teacher responses and from focus group notes, I led a discussion on the suggestions made for implementation process changes as well as changes to the questions from the data collected. The following suggestions from the questionnaires and focus group discussions were discussed.

- Allow teachers to determine or sign up for time period or window from end of February to first of May that worked around testing and individual subject schedules
- Shorten question list as well as revisit student capabilities to determine answers to particular questions.
- Require confidentiality rather than anonymity—fearing the lack of student trust this might initiate this idea was abandoned.
- Ensure that students take only once.It was decided that this could be facilitated by opening and closing the survey times and day according to teacher sign up lists.
- Give evaluations all at once. After discussion it was decided that it was still better to administer the evaluations on a teacher by teacher basis especially because the time period or window would be more spread out rather than making students take evaluations at one sitting and risk the 8th teacher not being given equal time or consideration due to student fatigue.

- Go from 7 to 0not 0 to 7. A suggestion from the social science department, “Trust us. It’s a psychology thing.” This approach, while not standard, was agreed upon as our local prerogative to “give it a try.”
- Condense the number of questions.
- Would students define “effective class time” the same or the word “professional” or “fair?” Eliminate or change wording of ambiguous questions.
- While initially considered, the consensus was this would add to an already time intensive process and could perhaps be addressed by principal led presentation before the evaluations on teaching students the importance of being specific on concerns or praise in their feedback in the open ended- questions. But because question one still had the ambiguous word of “effective” and understanding the student perceived why behind a low or high score on this question and additional comment box was placed just after this question.
- Because there would be a large number of students in the population, this could be done while still protecting student anonymity and would give teachers additional insight based on time periods or subjects or level of students in interpreting the data.

The faculty gave input and it was decided the question committee would reconvene and prepare a re-draft of the questions. Probably because there had been such a collaborative buy in to the first set of questions, the proposed revisions were minor even trying to accommodate the proposal to shorten the number of questions was difficult because of the attachment to the importance of all of the previous questions.

The same question committee from the first trial met again at a lunch meeting in January. The first question, "My teacher uses class time efficiently (no wasted time)," was one that reopened the question as to whether or not students could determine what was or what was not class time spent efficiently. The "no wasted time" clause had been added to clarify, but some teachers felt it again led to a negative connotation for the questions. So it was decided to leave the clarifying or leading phrase out.

In the attempt to condense the questions down to ten, it was decided that question number three and fourteen were similar. #3, I believe that my teacher wants me to succeed. #14, "My teacher sincerely cares about me as a person." But rather than just eliminate the idea behind question three the decision was made to combine the two questions to read, "My teacher sincerely cares about me as a person and wants me to succeed."

However, I had scheduled a meeting with the University of Utah Educational Policy Group, the group creating possible evaluation tools for the Utah State Office of Education, to discuss and inform our archival study. After that meeting, the idea of combining questions within one question was determined to be faulty practice. So the question was left at the one question "My teacher sincerely cares about me as a person," assuming that if a teacher did, then he or she would want the student to succeed as well.

Question 4, "My teacher is fair and consistent," was immediately rejected as being difficult for students to equally define what was "fair" or as ambiguous according to consistent in what? Question 6 was also thought to be difficult for students to determine. "My teacher acts like a professional." Do students know what it means to act like a professional? And sometimes being willing to go beyond the ordinary, perhaps a little unprofessional but not unethical, to get students excited should not be seen as a negative. So the question was eliminated.

Question 7, “I know what my teacher wants me to learn each class period”, seemed to be dictating the response of asking students to hold teachers responsible for stating a daily objective. However, the principal felt like this question was important in facilitating the school-wide goal of having all teachers have a clear purpose for every class period. Teachers agreed that this would be helpful to teachers and maybe even increase accountability. But the question also seemed similar to Question 8, “I am learning what I should be learning in this class.” This question was eliminated as in was decided this was a content area question that would be difficult for students to determine. So the Question 7 was left in and Question 8 eliminated.

Question 9 “My teacher encourages active student involvement and participation in the learning,” it was decided, was redundant and could be shortened to “My teacher encourages active student participation.”

Question 12, “My teacher is available for me to get help outside of class” was discussed as being different according to teacher’s individual schedules. Coaches or activity or club advisors might be penalized. And Question 2 addresses students receiving adequate help. So Question 12 was eliminated.

After debating on Question 13, “My teacher demonstrates good knowledge of the subject of this class.” It was decided that if students could not even believe that a teacher had good knowledge of the subject area that a major problem would exist. So it was left in.

In addition, it was brought up that we did not have a question regarding classroom management or control which was one of the characteristics that the MET Project (2010) determined as not only important by recognizable by students. So the addition of “My teacher manages classroom behavior” was added. With the revision of some questions, the addition of one question and the elimination of five, the list originally at 14 was now at the original goal of

10. The revised question list was approved by the committee and accepted as an informed improvement over the first (See Appendix B).

Next the committee re-examined the student self-evaluation questions and found that one of those questions had been left in the negative stance. If it was unfair for teachers to have the questions negatively worded, the same should go for the students. Question 3 was changed from "I rarely miss class," to "I attend class regularly" which students then would decide to what degree the positive statement was true or not.

The social science department wanted to try numbering the questions from the positive down to the negative which does not follow the ordered continuum of the typical Likert scale, but when dealing with adolescent patience in perhaps a tendency for choosing numbers the quickest and easiest way, perhaps it was worth a try. So the new scale would be re-ordered in the second implementation (see Appendix B) with 1 being the highest score and 7 being the lowest.

The teacher questionnaires Likert questions and open-ended questions were re-designed to specifically draw out information gained from comparisons from Phase I to Phase II. The focus group questions were the same again as the teacher questionnaires except designed for a response addressing the whole department rather than individual teachers.

- Determine the level of value found in student evaluations (Question 1 Overall: I found the student survey experience valuable),
- Determine level of changes made in instruction based on student feedback (Question 2: I made changes in my instructional practice based on student evaluations (slight wording change from will make to made changes because teachers are in the second phase),

- Identify level of reflection of teacher practice facilitated by student evaluations (Question 3 Student evaluations caused me to reflect on my teacher practice.),
- Identify level of teacher concerns over students ability to judge teacher performance (Question 4 I feel students are qualified to rate teacher instruction).

In addition, two new items were added to Phase II questionnaire:

- Determine level of reduction, if any, in levels of teacher anxiety over student evaluations having been through the process once (Question 5 My level of anxiety has been reduced after participating in the student evaluation process).
- Determining the degree to which collaboration and opportunities for input on the part of teachers reduced anxiety (Question 6 My level of anxiety was reduced by the collaborative process allowed in the implementation of student evaluations) (see Appendix C).

The questionnaires open ended questions remained the same except for a slight rewording to ensure that all questions are in an open-ended format and had the addition of questions specifically referring to the differences between the two rounds and the level of anxiety reduction in teachers (see Appendix C). Additional questions included the following:

- How did your evaluation results differ from last year?
- Identify changes you noticed due to question revisions from initial evaluations?
- What changes did you make in instruction this year based on last year's evaluations?
- What additional changes will you make in instruction because of the results of this year's student evaluations?

- What concerns were reduced by changes made in the implementation process from the initial to the second round?
- What ideas do you have on how to make additional changes to the implement process?

The data collected from Phase II teacher questionnaires also included basic demographic information on the participants, gender, and length of time in the teaching profession. The demographics allowed for further data analysis in Phase II.

Implementation Phase II. The new revised version of the student evaluation tool was approved by the whole faculty. And now all was in place for the process to take place again near the end of the year. As I had a new position as a principal in another school, the administration of the student evaluations fell to the principal and new assistant principals. I served as a consultant and active participant while no longer in the official role as an administrator.

Before the implementation process this time, as suggested in Phase I, the principal went on the school morning announcements and made a presentation to the student body as a whole regarding the value and purpose of student evaluations as an opportunity to give feedback that would be valuable to their teachers. Students were informed that crude remarks or personal attacks would be removed. The same student survey instrument and implementation process was used as in Phase I, freeing up some time, as questions only had to be slightly changed and students were familiar with the process. The principal and most of the faculty remained the same as well as the two-thirds of the student body, so the implementation process was easier to facilitate since it had been done before. The commitment level of participation for Phase II was 100% including the two teachers that had opted out before and all new faculty members.

Three portable Google Chromebook labs had been purchased and were expected to speed up the process of implementation. However, the delivery was delayed, so evaluations were not able to be administered as early as had been called for in Phase I. Teachers were given the option to have their students taken to the computer labs in April or wait until the first of May for the portable labs. Most teachers opted to wait for the Chromebook labs.

Near the end of the process, I came and spent a day working with the assistant principals to help speed up the process. I was able to quickly go in and out of classrooms with the portable labs, experiencing the reduction in time needed for the implementation made possible by this new technology. Students, too, obviously enjoyed using the new computers, and participated willingly and with the same enthusiasm noted in Phase I. With the delay in receiving the labs, a smaller percentage of classes were able to be surveyed in Phase II. But every teacher had at least two classes completed with additional evaluations being offered at teachers' requests. Surprisingly, there were almost 70% of teachers who wanted and received input from additional classes.

Nonarchival Phase III. The data collected from the archival studies led to the addition of a Phase III to test a theory that had emerged from the initial phases. The current faculty of the high school as well as teachers who participated in the initial phases were emailed a one question survey using Qualtrics as the survey instrument. The response rate was 39 teachers which was similar to the participation response for the teacher questionnaires from Phase I and II. This question represented a one-word change from a question from the Phase I teacher questionnaire. Phase I question asked, "Do you believe students are qualified to rate teacher instruction?" and Phase III question asked, "Do you believe students are qualified to give feedback on teacher instruction?" The survey question contained a definition of the word "rate" and "feedback" along

with the new question. Comparison between Phase I and Phase III teacher response to this question could then be added to the data analysis.

Methodology

The vastly differing contexts of school settings "cause scientists great trouble in trying to understand school life. . . In this hardest-to-do science, educators often need knowledge of the particular—the local" (Berliner, 2002, p. 19). This research looked at a federal issue from a *local* perspective. The U.S. Constitution left the control of education to the states and states passed much of that control over to *local* districts and *local* school boards. Yet, with the new globalization of economics, the focus has switched to the need for a *national* vision and standards. But, in the United States, national vision still is implemented through grassroots buy-in.

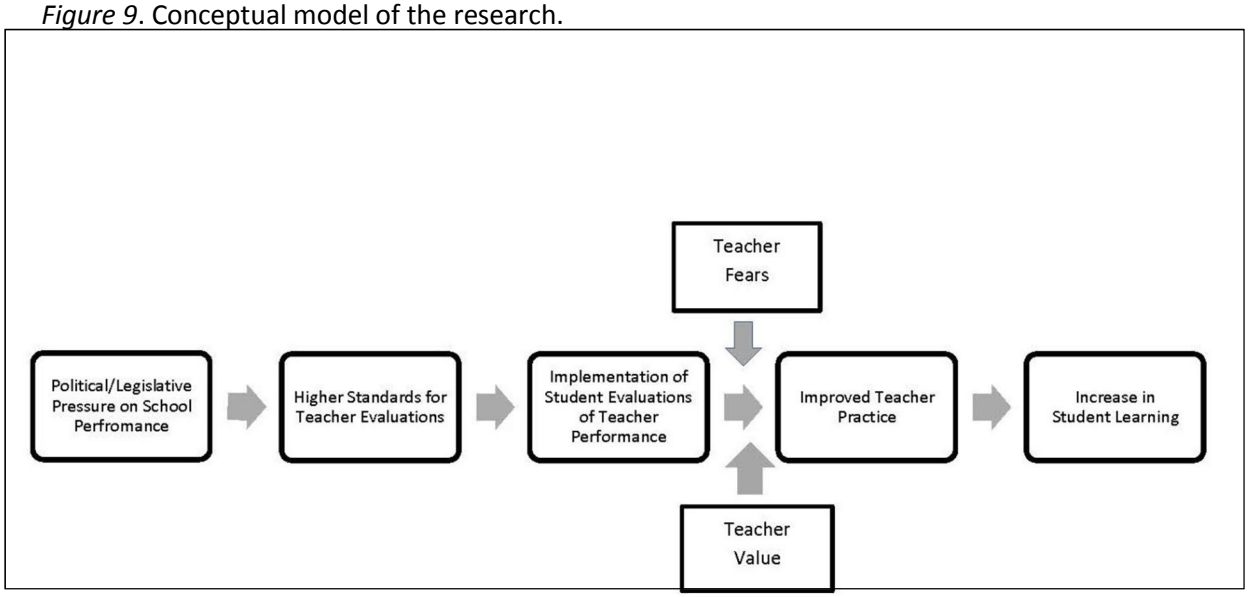
> Conservative Republicans, liberal Democrats, and organizations representing local educators are likely to throw up their hands and argue that school reform can't possibly come only from the outside, via external standards and accountability. They're right. Local educators and school boards have to buy into reform if it's to be successful, since they're the ones implementing it (Toch, 2012, p. 67).

Educational research is highly contextualized or situation driven. Every situation, every state, every district, every school, every teacher, and every student is unique. Case study research often uses qualitative and quantitative methods to examine a context-specific setting (Patton, 2001), a real-to-that-world setting. For this research, that setting was the implementation of student evaluations of teacher performance for the first time at one specific high school in the state of Utah. The focus of the research would not be on the reliability of the actual student input, but rather to view the process from the perspective of the teachers, viewing their anxiety level

and their reactions as to whether or not they found value in the student evaluations and their suggestions for improving the implementation process.

Mix-method research can provide a more descriptive account, "rich" (Gibbs, 2007, p. 21), or "thick" (Mason, 2002, p. 3), of what student evaluation of teacher performance specifically looks like, maybe even feels like, for teachers in this school. Mixed-methods research allowed the researcher to seek illumination among such uncertainty surrounding student evaluations. The conceptual model in Figure 9 illustrates the major constructs of student evaluations to be explored as explained in the literature within this specific school environment.

Figure 9. Conceptual model of the research.



The study used quantitative statistical analysis to provide greater confidence in the transferability (Creswell, 2002) of the qualitative research data. Thus, the researcher employed "triangulation mixed method design to simultaneously collect both quantitative and qualitative data" (p. 564), and used the results to describe and explain the research questions "blending the strengths and neutralizing the weaknesses of each method" (p. 561).

The research accessed both an emic and etic approach. Because the study lies within the culture of one school, this lead to an emic view. While the researcher may have pre-disposed theories based on the literature, those prior ideas did not override or inhibit the participants from telling their own story within their environment. "Emic constructs are accounts, descriptions, and analyses expressed in terms of the conceptual schemes and categories regarded as meaningful and appropriate by the native members of the culture whose beliefs and behaviors are being studied" (Lett, 1990, p. 130). In taking an emic approach, the researcher tried to put aside prior theories and assumptions so that the data created themes, patterns, and concepts as they emerged.

At the same time, this research used the etic approach that allowed for comparisons of outside resources and viewpoints to those of the teachers within this school's culture. A completely etic approach risks missing potential new ideas. At the same time, since all researchers come with previous theories, it may be impossible to be purely emic. Given that the researcher had been part of the culture being studied and also participated in a thorough review of viewpoints outside of the culture, both a natural tension as well as ease exists between balancing the two approaches. Being able to openly embrace both lenses is an advantage to the researcher to look at the study from both the perspectives of an insider and outsider.

Thus, mixed-methods methodology allowed for the telling of the unique perspective of one school. But, the study also allows for the drawing of comparisons from that school perspective to the body of literature, "explaining what people and situations have in common and doing this with reference to existing theories and concepts" (Gibbs, 2007, p. 26).

School Context

The researcher knew the context, and the context knew the researcher, making the data more valid because of an existing trust relationship. In fact, Spindler and Spindler (1987), in their

criteria for good qualitative ethnographies, felt that observations must be contextualized, both in the setting of the study and in further contexts beyond that setting, as relevant. While the context is known, the process of the study was unknown to both the researcher and the participants, thus the researcher was able to assume an observer participant role. As an assistant principal, the researcher facilitated or implemented the evaluations but as a researcher analyzed and observed the process, becoming a participant-observer. Spradley (1980) suggests that the participant observer has two goals: (1) to engage in activities appropriate to the situation and (2) to observe the activities, people, and physical aspects of the situation (p. 54). Researchers in this framework are therefore alternating between the participant and the observer roles, yet experiencing both simultaneously (Spradley, 1980, p. 57). The researcher choose this high school for the study because the school was ready to go through the student evaluation process and because the researcher would have access both to the school and data from the implementation process. The potential data would be mutually beneficial to the school and to the researcher both as a principal and as a researcher**.** The GAP analysis in Figure 10 illustrates the current "Where are we now" context for the school used in the archival study before the initial implementation of student evaluations.

STRATEGIC OBJECTIVE	CURRENT STANDING	DEFICIENCY	ACTION PLAN
Prepare for upcoming legislative mandates from Utah HB 64 for multiple perspectives in teacher evaluations	Traditional principal yearly observation used as teacher evaluation	Yearly teacher evaluations use only one perspective	• Assemble multiple stakeholder student evaluation implementation committee Winter 2012 • Research and create questions for student evaluation tool in committee • Present plan to faculty • Re-assess plan according to faculty input • Implement initial trial of school-wide student evaluation process through collaborative input Spring 2012 • Facilitate faculty reflection through questionnaires and focus group discussions on initial trial • Collect data • Present results to faculty • Reconvene stakeholder committee for re-evaluation of plan Winter 2013 • Present new plan to faculty • Modify plan according to stakeholder and faculty input Spring 2013 • Implement 2nd trial of student evaluation process • Facilitate reflection again through questionnaires and focus group discussions • Present data to faculty • Reconvene stakeholder committee for re-evaluation • Make recommendations to faculty for standard practice for administration of student evaluations • Request school plan for adoption of Utah State School Board to be used in compliance with HB 64 as additional perspective in teacher evaluations
Expand student voice opportunities to all students	Student voice opportunities have been priority: student council, youth school board members, club presidents (Hawks Nest), fan club, athletic captain's table	Student voice opportunities limited to school leaders - not accessing all students	
Inform teacher practice through standard practice of accessing student feedback	Solicitation of regular student feedback or evaluations limited to only 5% of the faculty	90 % of teachers not using consistent method to obtain student feedback from all students	
Increase student learning and engagement	Mediocre/Middle end of year CRT scores, AP scores, CTE scores, passing rate, high graduation rate	Test scores and passing rate not equal to demographic capacity Students not graduating with necessary high level skills	

Figure 10. Preresearch GAP analysis of high school.

This high school would be typical for other Utah high schools and schools around the nation in that it follows what the literature deems the pattern for current teacher evaluations requiring only one perspective, the annual principal observation. In preparation for the requirements of Utah HB 64 by the 2015-16 school year for multiple perspectives, one being student evaluations, the school hoped to use the interim time period to prepare for change through a collaborative process that allowed for the creation of an individualized plan. This research looked at not only the opportunity, but the mandate to include student voice.

Anticipating teacher anxiety under the higher stakes ahead as the evaluations will eventually be used for a yet to be determined teacher ranking or grading system and linkage to salaries, the goal of the trial plan was to decrease teacher concern by allowing them ownership and input into the implementation process as well as time to reflect on the value of the feedback, perhaps making improvements to instruction that will facilitate better student feedback in the future high stakes setting. While the catalyst for the plan was the legislative mandate, the vision was to decrease teacher anxiety and optimize the benefits of increasing student voice as a vehicle to improve teacher practice and increase student learning.

To alleviate teacher anxiety and facilitate willingness to participate in the research, the evaluation statements from the actual student feedback from both Phase I and Phase II are not included in the research data. Teachers were assured that the student evaluation data during both phases will only be seen by themselves and the school principal. The researcher did not view the actual student evaluations because the research focused only on the teachers' responses to those student evaluations. This assured teachers that no publication of student evaluation results would occur, preventing researcher bias, and maintaining the focus of the study on the perspective of teachers.

Target Population

The researcher gathered archival data from the perspective of teachers on the implementation of student evaluations from the target high school principal. The target population of teachers was from one Utah high school located in a suburban area of Utah County. These participants were not systematically selected, thus their selection represents a form of purposeful sampling which may or may not be representative of the population of teachers as a whole. Because the researcher had access and trust with the selected and willing participants, this school was chosen for this study. The school also represented a school and faculty facing the impending implementation of student evaluations. Thus, anxiety and a natural vested interest and sense of urgency existed within this school environment. The strength of this sample would be that it represents almost an entire group or faculty with only two no non-responders (40 out of 42) total faculty members in Phase I, 38 in Phase II, and 39 in Phase III (full and part-time teachers of the high school). With the ability to describe in detail the experience of a small group (Creswell, 2002), this study—while unique to its own context—was able to examine a common problem that each school must solve within its own context.

In the archival data from the two-year implementation of student evaluations of teachers, teachers from eleven different subject departments were invited to participate: Business Marketing, English, Family and Consumer Science, Fine Arts, Foreign Language, Health Life Styles, Mathematics, Science, Social Sciences, Special Education, and Trade and Technical. The final faculty population of 51 teachers also included three teachers that were added after the initial program study. The faculty demographics included 32 veterans (over five years) (63%) and 19 new (less than five years) teachers (37%) as well as 31 males (61%) and 20 females (39%). Prior to accessing the data for this study, these faculty participants and school leaders

signed an IRB-approved informed individual consent form given that the data was not be de-identified (see Appendix A).

The high school was a fairly new high school. It was 4A school with designations ranging from 1A to 5A based on number of students enrolled, putting the population size above average for Utah high schools with an approximate student population of 1133. The Utah average enrollment for public high schools is 801 students according to the National Center for Educational Statistics (2012). Demographics of the school (US News/Education, 2012) showed a low degree of ethnic diversity, 1%, but some diversity in family income level with 25% of the student body being economically disadvantaged determined by free and reduced lunch. The school test scores at the end of the 2nd implementation showed 91% proficiency in English and 29% proficiency in math compared to the state average of 81% in English and 36% in math. The school's 95% graduation rate was one of the highest in the state of Utah which is 79% overall.

The researcher believed that the opportunity for teachers to inform and contribute to the process of evaluating teachers, which is almost a future inevitability, provided an incentive for and facilitated a high degree of faculty participation in the research sample, reducing coverage error, sampling, and non-response error (Creswell, 2002).

Data Collection

Archival data was collected from the sample of participating faculty to study or discover "attitudes, opinions, behaviors, or characteristics" (Creswell, 2002, p. 60) of the school's teachers facing and experiencing student evaluations as part of evaluating their performance. The data collection consisted of accessing the archival data which included the results from both phases of teacher questionnaires and focus group departmental discussions conducted with the use of a question guide to facilitate elaboration of questionnaire responses.

Focus groups are ideal for mixed-method research in the very fact that they call for multiple perspectives of a reality Focus group discussions have definite benefits, facilitating group interaction, encouraging research participants to explore the issues in their own words, generating their own questions, and adhering to their own priorities and points of view, taking the research in new and often unexpected directions (Sinagub, Vaughn, & Schumn, 1996). In fact researchers have felt that group discussions "reveal more critical comments than do one-on-one interviews" (Kitzinger, 1995, p. 299). The use of open-ended questions in the focus group discussions also allowed participants to share how the participants interpreted their experiences and allowed the focus group leader to ask follow-up questions to expand on or clarify responses.

The focus groups were conducted by department—groups that "naturally occur" (Kittzenger, 1995, p. 311) within the school environment. Focus groups notes were not recorded because of expressed concerns from teachers and given that this practice may defeat the purpose of the process which is seeking to decrease teacher anxiety.

The data collected included quantitative and qualitative data from teacher questionnaires and qualitative data from focus group discussions from both phases. Data sources, Likert response and open-ended questions from teacher questionnaires and focus group notes, as shown in Table 1, from both phases informed the research questions.

Table 1

Archival Data Sources

Phase I 2012	Phase II 2013
Teacher questionnaires: Open-ended questions and agree/disagree Likert response questions Appendix C	Teacher questionnaires: Open-ended questions and agree/disagree Likert response questions Appendix E
Department focus group interviews: Open-ended questions Appendix D	Department-head focus group interview: Open-ended questions Appendix F

The research questions for this study were aligned to match the teacher questionnaire items and focus group questions from Phase I and Phase II. A question from the questionnaires or focus groups may address more than one research question. Table 2 shows this alignment.

Table 2

Research Questions Aligned to Teacher Questionnaire and Focus Group Questions

Phase I questionnaire Likert items	Aligned research questions
Overall, I found the student survey experience valuable.	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?
I made changes in my instruction due to student input from the survey.	Will teachers change or alter instructional practices based on the input of their students?
Student evaluations caused me to reflect on my teacher practice.	Will teachers reflect on their practice based on student comments? To what extent?
I believe students are qualified to rate my instruction.	Do teachers feel comfortable about student ability to judge instruction? To what extent?
My level of anxiety has been reduced after participating in the student evaluation process.	What can be done to improve the implementation of SETs? Will teachers find going through the process helpful? To what extent?
My level of anxiety was reduced by the collaborative process allowed in the implementation of student evaluations.	What can be done to improve the implementation of SETs? Will teacher input and ownership in a collaborative process alleviate some teacher anxiety over student evaluations? To what extent?
How did your predictions differ from actual results?	Will teacher input and ownership in a collaborative process alleviate some teacher anxiety over student evaluations?
What insights did you gain from the data? Any information valuable from comparing lower scores to higher scores?	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?

Phase I questionnaire Likert items	Aligned research questions
Did you gain any valuable information from student comments section? Explain?	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?
Now that you have seen the results, would you like to add or take out any questions?	Will teacher input and ownership in a collaborative process alleviate some teacher anxiety over student evaluations? To what extent?
Will you make any changes in instruction because of the results of your student evaluations? What specific change will you make?	Will teachers change or alter instructional practices based on the input of their students? Specifically what changes?
Overall what did you find the most valuable from the student evaluations?	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?
What are your biggest concerns with student evaluations?	Will teacher input and ownership in a collaborative process alleviate some teacher anxiety over student evaluations? To what extent?
Do you have ideas on how to implement or administer student evaluations other than the way these were done?	Will teacher input and ownership in a collaborative process alleviate some teacher anxiety over student evaluations? To what extent?
Overall, do you see a value in student evaluations?	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?
Any other information or comments	Will teacher input and ownership in a collaborative process alleviate some teacher anxiety over student
Overall, I found the student survey experience valuable.	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?
I made changes in my instruction due to student input from the survey.	Will teachers change or alter instructional practices based on the input of their students? Specifically what changes?
I used input from student evaluations last year to facilitate classroom discussions this year	Will teachers change or alter instructional practices based on the input of their students? Specifically what changes?
Phase I questionnaire Likert items	Aligned research questions
Overall, student evaluations have caused me to reflect on my teacher practice.	Will teachers reflect on their practice based on student comments? To what extent?
My level of anxiety has been reduced after participating in the student evaluation process.	What can be done to improve the implementation of SETs? Will teachers find going through the process helpful? To what extent?
The collaborative nature allowed in the implementation of student evaluations increased my level of support for the process.	Will teacher input and ownership in a collaborative process alleviate some teacher anxiety over student evaluations? To what extent?
How did your evaluation results differ from last year?	Will teachers reflect on their practice based on student comments? To what extent?
Identify changes you noticed due to question revisions from initial evaluations?	What can be done to improve the implementation of SETs? Will teachers find going through the process helpful? To what extent?

Phase I questionnaire Likert items	Aligned research questions
List specific changes made in instruction this year based on last year's evaluations.	Will teachers change or alter instructional practices based on the input of their students? Specifically what changes?
What changes did you make in instruction this year based on last year's evaluations?	Will teachers change or alter instructional practices based on the input of their students? Specifically what changes?
What additional changes will you make in instruction because of the results on this year's student evaluations?	Will teachers change or alter instructional practices based on the input of their students? Specifically what changes?
What did you find the most valuable from the student evaluations (either time)?	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?
What did you find the least valuable from the student evaluations (either time)?	Will teachers find value in the feedback of their students despite possible negative or hurtful comments? To what extent?
What did you notice by not having as many classes give feedback this year?	What can be done to improve the implementation of SETs? Will teachers find going through the process helpful? To what extent?
What ideas do you have on how to make additional changes to the implementation process?	What can be done to improve the implementation of SETs? Will teachers find going through the process helpful? To what extent?
Any other questions or comments	What can be done to improve the implementation of SETs? Will teachers find going through the process helpful? To what extent?

Note. Phase II Questionnaire Open-Ended Questions and Department Focus Group Questions were the same with the addition of the word department to each question.

Establishing trust. In sharing personal heartaches and heartwarming self-administered student evaluation moments, the researcher hoped that to establish trust with the faculty. While an administrator, the researcher was one of them, first and foremost and always a teacher, who not only understands, but has lived and felt the pain and power of student evaluations, a trust that would be crucial in going forward with this research study.

While this high school faculty would be considered typical in make-up, in reality, they are somewhat unique. The school was new and the faculty, and all stakeholders, had participated in the collaborative venture of creating culture and tradition. With deliberate and purposeful measures that involved all stakeholders, including and most especially the students, the first

principal set about to design a distinctive sense of ownership in who and what the school became together.

And yet, even within this remarkable atmosphere, the thought of student evaluations, the passage of House Bill 64, caused fear and trepidation among *this* faculty and not just for a few teachers, but for all. This anxiety level is significant in grasping the full impact or level of anxiety that teachers, even the best of teachers, feel in opening themselves up to even the possibility of criticism from their students.

However, despite the faculty concerns, having gained their trust as a researcher but also a faculty member, the researcher presented on what legislatively lay ahead and on the potential power in giving students voice. Overwhelmingly, the faculty as a whole supported the idea of a school wide trial run of implementing student evaluations. Part-time and full-time teachers, even the police officer who teaches one class, were willing to take part. Only two teachers later privately chose to opt out, both close to retirement, making 49 teachers out of 51 participating in allowing students to evaluate their classes.

The principal did say that he would like them all to participate, giving his vote to the value of the project, but that he would not force anyone this first time. He is a trusted, beloved principal and his request carried weight. Also, the researcher assured the teachers that this would be a no stakes or low stakes trial run. For research purposes, the researcher was interested in their feedback about their students' evaluations and the process itself. Informing the process and getting a chance to have a say in the implementation was also an incentive. The researcher promised that not to even look at the results of the actual student evaluations, so there could not possibly be anything published about student comments. Again, there was a relationship of high trust, them with the researcher and the researcher with them.

Teachers expressed some similar concerns over the issues reflected in the literature, personality concerns, time of day of the class, number in the class, angry or troubled students just taking the opportunity to vent, and being the eighth teacher a student evaluates vs. the first. No concerns, at least expressed openly, were noted over gender, age, appearance or ethnic background. However, this faculty would not be considered ethnically diverse. This issue of elective teachers vs. core teachers was a particular point of discussion. Students choose to take electives because of an interest but are required to take math or English. So the fear was that elective teachers would score better. It is interesting that the term better or higher was used when there was going to be no comparisons drawn. The individual teacher and the principal, it was determined, would be the only ones to see the results. There would be no school wide tallies of scores on questions. The feedback was to be for individual teacher use.

The only requirement from the principal was that teachers were to view their evaluations and set one goal to work on based on the feedback. In the end of year exit interview, when check-out sheets are turned in, teachers were to come prepared to discuss his or her goal with the principal. The request from the researcher was to participate in an anonymous teacher questionnaire and in a department focus group discussion reflecting on the value and use of the feedback and the implementation process, offering recommendations for the second trial run the following year, giving two trials before the legislative mandate is scheduled to go into effect.

Teacher questionnaire archival data collection for Phase I. Student survey results were immediately made available to teachers, giving teachers time to review their results before filling out the questionnaires and participating in group discussions Teachers were given hard copies, paper and pencil, of individual teacher reflection questionnaires at faculty meeting to be returned to the researcher's school box. An email was sent out to remind all teachers to make

sure they had returned the questionnaires. The questionnaires were to be anonymous and included four agree/disagree Likert response questions for comparison purposes and open-ended prompt questions. Questions asked if teachers found value in the student evaluations and to what degree, if they would make changes in instruction, if evaluations caused reflection, and if they believed in student capabilities to make judgment on instruction as well as open ended questions asking for feedback on the implementation process itself, suggested changes in the process, and further reflection on value of feedback. The participation response rate on teacher questionnaires was 40 out of 51 teachers or 78%.

Teacher questionnaire data archival collection for Phase II. Survey results were immediately made available to teachers, giving teachers time to review their results before filling out the questionnaires and participating in group discussions as in Phase I. Questionnaires were given out to all faculty participants at the last faculty meeting of the year. Teacher questionnaires in Phase II used the same first four questions and the same Likert scale. The Phase II questionnaire also included the addition of two more agree/disagree statements to address the level of anxiety reduction in having participated in the student evaluation process and in being given the opportunity to be part of the collaborative process. The questionnaire was administered in the same way as in Phase I so that any differences in findings are not attributable to differences in how the data was collected with the exception of teachers returning the questionnaires to a box in the front office facilitated by the office secretary. The teacher participant response rate on the questionnaires was 38 out of 42 or 90%.

Focus group archival data collection for Phase I. In Phase I, each department discussed the focus group questions in the regularly scheduled PLC meetings, allowing for maximal teacher attendance and department participation. Participation in focus groups was 88%

with some teachers being gone to other assignments during collaboration PLC time the week of focus meetings. Since these meetings were occurring simultaneously, the researcher could not attend all of the meetings. The researcher had met with department heads before the focus group discussions to establish a common protocol and answer questions about the process. Department heads facilitated the group discussion on questions similar to the individual teacher questionnaires and recorded notes. The department heads took notes by hand directly onto the focus group question document and turned a hard copy into the researcher directly after the meetings.

Focus group archival data collection for Phase II. Focus group discussions from Phase II included a whole faculty discussion during the end of year faculty meeting during early out collaboration PLC time. No predetermined questions were used. Teachers were allowed just to freely comment on the process and draw comparisons between the two years of implementation. The principal recorded notes on his laptop, allowing the researcher to conduct as well as participate and facilitate a discussion. An electronic copy was sent to the researcher. Additionally, a lunch time focus group discussion met the following Monday. This time the focus group consisted of department heads, not whole departments, where the principal again took electronic notes. Each department was represented and thus 100% attendance. The focus group questions were the same as those on the Phase II teacher questionnaire with questions calling for the discussion of the department as a whole rather than individual teachers and also calling for comparisons between the two phases (see Appendix D).

Additionally, in Phase II, another focus group included the researcher meeting with the current administration by appointment to facilitate an administrative focus group discussion. This discussion specifically noted changes noted by administrators between the two different phases

of the implementation process. The researcher took detailed notes during this discussion. Teachers had been required to bring a goal set based on a reflection from the student evaluation results as part of the end of year check-out process with the principal which gave the principal additional knowledge of teacher reactions to the process.

Nonarchival data collection for Phase III. The data collected from the archival studies led to the addition of a Phase III to test a theory that had emerged from the initial phases. The current faculty of the high school as well as teachers who participated in the initial phases were emailed a one-question survey using Qualtrics as the survey instrument. The response rate was 39 teachers or 93% which was similar to the participation response for the teacher questionnaires from Phase I and II. This question represented a one-word change from a question from the Phase I teacher questionnaire. Phase I question asked, "Do you believe students are qualified to rate teacher instruction?" and Phase III question asked, "Do you believe students are qualified to give feedback on teacher instruction?" The survey question contained a definition of the word "rate" and "feedback" along with the new question. Comparison between Phase I and Phase III teacher response to this question was then added to the data analysis.

Data Analysis

After the data were collected, the data were analyzed to discover patterns and themes that emerged from the data. By comparing the results of the two phases and an additional third phase as well as comparing the results to the literature, the data analysis allowed for interpretation of findings as it related to the research questions. Qualitative and quantitative analysis was done separately and then combined and linked in Phase II to create a detailed picture of the teachers' perspectives of the implementation of student evaluations.

With the use of a small sample, case study, the extent to which this sample represents teachers beyond this school cannot be known and thus use of statistical analyses was somewhat limited (Creswell, 2002). However, analyses can ascertain that certain characteristics existed within this study sample. The traits of particular demographic sub-groups within this context can also be described in an exploratory way.

Triangulation. In one sense, validity in research is very specific to the test to which it is applied (Patton, 2001). However, triangulation using multiple data sources provides multiple perspectives on a single reality (Healy & Perry, 2000). This research used triangulation and mixed method design of both qualitative and quantitative methods collecting individual teacher questionnaires with open-ended survey items, close-ended agree/disagree statements based on a Likert scale, and two sets of focus group discussions.

Triangulation facilitates corroboration, not to confirm the accuracy of a reflection, but rather to help ensure that the data accurately reflects the results, whatever they may be. The purpose of corroboration in this research was to help ensure credibility and documentable evidence (Stainback & Stainback, 1988), regardless of the outcome. The final report of the research findings will communicate clearly the data analysis methods used for the findings, allowing the intended audience, other public schools, the opportunity for adaptation or application. Credibility and transferability rather than replicability provide the lenses of evaluating the findings of this mixed method research (Hoepf, 1997).

The individuals and groups who supplied the data were themselves part of the interpretation. “Credibility needs to be established with the individuals and groups who have supplied the data for inquiry” (Erlandsen, 1993, p. 30). This credibility standard will also foster confirmability in that the findings were open to the public scrutiny of the participants which

aided in protecting the product from the bias of the researcher (p. 34). Dependability was established by repeating the implementation of student evaluations, Phase I and Phase II, to insure the "criterion of consistency" (p. 32) within the same respondents and setting. Although no new situation will match exactly the original study, major conditions may be similar.

Given the theoretical perspective of the original researcher and following the same general rules for data collection and analysis, plus similar conditions, another investigator should be able to arrive at the same general scheme. The discrepancies that arise should be resolvable through re-examining the data and identifying special conditions operating in each case (Corbin & Strauss, 1990, p. 15).

Grounded theory design. The process of evaluating teacher performance is at the forefront of educational reform. How teachers are evaluated is a process. The question is how to use student evaluations as a piece of that process "A grounded theory research design is a systematic, qualitative procedure used to generate a theory that explains at a broad conceptual level, a process, an action, or interaction about a substantive topic... this theory is a process theory—it explains an educational process" (Creswell, 2002, p. 439). Discovering some ways to effectively administer the process of student evaluations to access student voice in a way that reduces teacher anxiety, informs teacher instruction, and facilitates teacher reflection not dejection is the hope of this research.

The purpose of using the ground theory design was, first, to ensure that the findings were grounded in the data and, second, to generate a theoretical understanding about the implementation of student evaluation process based on the data representing the perspectives of the participants—the open-ended items from the teacher questionaries' and the focus group discussions. The purpose of this study was to "discover emerging ideas and to better understand"

(Creswell, 2002, p. 454) the use of student evaluations by examining the two-phase process of implementation at one high school for the first time. Glaser, who along with Strauss created the grounded theory approach, stressed the importance of letting the theory emerge through the data (Glaser, 1992).

The data collected on the teacher's perspective guided the research. Data analysis of the archival data included a continuous interaction and comparison between the first and second phases of student evaluations. Grounded theory seeks to not only uncover relevant positions, but also to determine how the actors respond to changing conditions. The researcher was responsible for catching the interplay between the two phases (Corbin & Strauss, 1990), an active participant trying to determine if the method used for administering evaluations can be refined through collaborative efforts to lessen teacher concerns while maximizing student voice.

Charmaz (1990) advocates a constructivist approach to grounded theory where the researcher could be viewed as a participant with a role in bringing questions to the data derived from values and experiences which are added to the feelings of the other participants as they experience the process mutually. Having been a high school teacher who created and used student evaluations and a university adjunct professor evaluated by students for the past eleven years, the researcher brought shared experience as well as a shared empathy to the discussion table. The data analysis exposed new understandings for both the researcher and the participants as interaction informed new ideas.

While evidence for the need to include students in school reform and some evidence of successful implementation of student evaluations has been presented in the review of literature, the final suggestions and conclusions of this research do not depend on theoretical models already developed. Within this unique context, a modified or personalized model of

implementation for this school emerged from within the data. Thus, the findings are applicable directly to the school in which the program study was conducted. Yet, "insofar as theory that is developed through this methodology is able to specify consequences and their related conditions, the theorists can claim predictability for it, in the limited sense that if elsewhere approximately similar conditions apply, then approximately similar consequences should occur" (Strauss & Corbin, 1998, p. 169). The ideas generated by this study may thus be transferable or have some relevance to similar environments and possibly for teachers, schools, and states interested in using this study as a springboard to inform or develop their own model of implementing student evaluations.

Qualitative data analysis. The open-ended items on the individual teacher questionnaires and the questions in the focus group discussions allowed participant teachers to elaborate or create their own responses, use their own words, according to their own experience. "The open question often elicits responses of items or ideas overlooked by the researcher; it uncovers themes" (Creswell, 2002, p. 406). Open-ended responses on the teacher questionnaires and notes from focus group discussions from Phase I and Phase II provided the researcher with "shared understandings" (Creswell, 2002, p. 206).

The analysis of the open-ended questions on the teacher questionnaires and from the focus group discussions followed the first three basic steps presented by Boeije for the Constant Comparative Method, "comparison within a single interview, comparison between interviews within the same group, comparison of interviews from different groups" (2002, p. 395). Research analysis employed the basic framework of grounded theory for qualitative analysis proposed by Marshall and Rossman (1999): organizing the data, coding the data (open, axial, selective), generating themes, testing emergent understandings, searching for alternative explanations.

In the initial open coding, similar codes aggregate together to form themes or main ideas or sometimes referred to as "categories, dimensions, issues, or perspectives" (Creswell, 2002, p. 267). These themes may be codes "more frequently discuss by participants, are unique or surprising, have the most evidence to support them, or those that might be expected" (p. 267). Open coding explores and identifies similarities, differences, and connections between the participant responses (Miles & Huberman, 1994). Open coding methods help guard against the use of preconceived notions of the researcher when confronting the data (Corbin & Strauss, 1990).

Secondly, grounded theory calls for a second tier, axial coding, where categories are related to subcategories and additional categories may be discovered. Finding and identifying patterns or regularities gives order to the qualitative data. Data is examined for patterns and for an understanding of where the patterns were not apparent (Corbin & Strauss, 1990).

Thirdly, selective coding or selecting the themes and patterns that emerged as the central or core phenomenon and the factors that influenced and actions, consequences and outcomes that resulted from these themes and patterns were identified. Identification of key themes and patterns facilitates the writing of the report or story, of the inter-relationships between the patterns identified in the axial coding and highlights theoretical explanations (Corbin & Strauss, 1990).

Qualitative data coding for Phase I. The Phase I open-ended responses from the teacher questionnaires and focus group discussions were initially analyzed individually by hand, manually. In open coding, each teacher response question by question was assigned a category in the form of marginal notes on each teacher questionnaire and focus note sheet. The categories were emergent as the data was reviewed. The researcher then noted repeated ideas, words, or phrases within the categories, categories that related to themes from the literature, and unique or

surprising findings. Thinking of the process as filing, placing each response in a file folder, the categories then were compared and combined and inserted into an Excel file spreadsheet listing the emerging categories from each question with a column that also recorded a quote that was representative of the category.

Axial coding was aided by sorting within Excel. Categories were combined, expanded or collapsed, reworded, and revised by evaluating patterns and reevaluated to assure that the data was representative of all teacher responses. The following list of reflective questions was used as a reference for this categorization process ("Tips and Tools: Coding Qualitative Data," 2012):

- What is this saying? What does it represent?
- What is this an example of?
- What do I see is going on here?
- What is happening?
- What kind of events are at issue here?
- What is trying to be conveyed?

Finally, selective coding methods were used as the Excel file was refined and collapsed to represent core or central themes that had emerged from the categories. Themes represented patterns or meaning from the data that specifically related to the research questions. Most themes came from repeated or high frequency teacher responses, but even if a theme did not come from an often repeated response that did not necessarily mean that the theme was not important. Subtle phrases or connections between responses or categories also revealed themes. The degree to which the theme added to the richness of or complexity of the research was another criteria used for selection. The researcher's judgment and experience as a teacher and administrator was also a key tool in selecting relevant themes. The researcher was also careful to avoid personal bias and

not go beyond the intent of the teacher responses so that the final analysis represented a comprehensive, but also meaningful view of the actual data (Creswell, 2002). The representation of the data would also be presented to the participants as a source of member checking to ensure credibility.

Qualitative data coding for Phase II. Phase II coding of qualitative data from individual teacher responses and focus group discussions was analyzed by repeating the coding steps of Phase I. This analysis additionally documented any new categories that emerged and also included categories for the new questions noting significant changes or differences in teacher actions as well as comparisons between the two phases. An Excel spreadsheet was created with the new categories. Quotes from this phase were added to represent the new categories.

Once the Excel spreadsheet for Phase II was uploaded in R software the central themes from the questions were compared collectively from both Phase I and Phase II. Differences or similarities between the two phases became new categories. The categories were refined and combined into core themes for Phase II. Theme selection included the same process as outlined in Phase I. But Phase I also acted as a starting point from which to build from for Phase II. Similarities and differences between the phases served as another frame of reference for selected themes. The Excel spreadsheet crafted the categories into core themes for this phase. This process revealed new questions as new themes emerged from the data which led to the creation of an additional Phase III to explore a new theme.

Quantitative data analysis for Phase I. In Phase I, the quantitative analysis examined the Likert scale questions from the teacher questionnaires. Teacher participants had indicated on the teacher questionnaires their level of agreement with the given statements by way of an

ordinal Likert scale. Likert response questions were quantitatively analyzed using R software to provide descriptive statistics.

R software was used to run basic statistical techniques to explore the frequencies of responses and frequency distribution of each question. Frequency of response was used to determine a mean and variance from each question count which could be interpreted for each question and used for comparison between questions. Bar graphs were created to represent the frequency of response for each question. A table was created to represent the statistics for each question. By examining the table for significant comparisons between questions, the researcher additionally created bar graphs that compared the variance and frequency of response between some questions.

Quantitative data analysis for Phase II. While the extensive qualitative analysis of Phase II might still be considered to be subjective in nature, this could be balanced with the more objective quantitative analysis. The quantitative analysis in Phase II examined the same Likert scale questions from the teacher questionnaires in Phase I with the addition of two more questions calling for teachers to make comparisons between the Phases. These questions were again quantitatively analyzed using R software to provide descriptive statistics.

R software was used to run basic statistical techniques to explore the frequencies of responses and frequency distribution of each question in Phase II. Frequency of response was used to determine a mean and variance from each question count which could be interpreted for each question and used for comparison between questions and additionally between Phases. Bar graphs were created to represent the frequency of response for each question. A table was created to represent the statistics for each question.

Additional Phase III quantitative analysis. Based on the analysis of Phase I and Phase II, a Phase III was designed. One additional teacher questionnaire was added, returning to the sample population to obtain further clarification on a possible key revelation. Because teachers' anxiety level was observed to still be high prior to Phase II, the researcher did not repeat the question in Phase II asking teachers to indicate if they believed students were qualified to rate teacher instruction. No indication appeared to exist that anything had changed that would influence a difference in response for this question. However, through the data analysis process, the researcher discovered a new possibility. In the data analysis, a particular comment from a teacher stood out. On one teacher reflection questionnaire with an overall negative response to all questions about student evaluations, the teacher had written in the additional ideas or concerns section, "I *actually* really like student feedback." The teacher was saying that he liked feedback, but, as indicated by his low responses on the questionnaire, disliked and did not value student evaluations. How could this be? Somehow, this teacher did not view student evaluations as feedback.

In looking at the question asked in Phase I, the word *rate* suddenly stood out. The teachers had been asked if they believed students were qualified to *rate* teacher instruction in Question 4. Is the purpose of student input or student voice to ask students to rate or evaluate teacher instruction or is it to give feedback on teacher instruction? Would teachers find the question "Do you believe students are qualified to give teachers feedback on instruction?" more acceptable, receive a more positive response, than the original question? The researcher discussed this thought with the high school principal at the research site and both agreed this was a distinct possibility or at least an important question to ask.

Thus, a Phase III survey on Qualtrics software was emailed out to the faculty. The survey contained a definition of the word *rate* and the word *feedback* and a copy of the previous question and the new question, asking teachers to note the wording change in the new question and to indicate their agreement level on the new question (39 teachers responded).

Response frequency graphs for this new question and the original Phase I enabled a comparison of responses based on mean, variance, and standard deviation. This time standard deviation and variance was used as the results were found to be statistically significant and perhaps would suggest some generalization to the population of teachers as a whole. The researcher used a two-sample *t*-test to see whether there was a statistically significant difference in the responses to the questions from Phase I to Phase III. A *t*-test is a parametric test used to determine whether two samples are significantly different from each other. A second verification was done using the Mann-Whitney U test. This test is a widely used non-parametric test that is often more powerful and efficient when the underlying distribution of the data is unknown.

Mixed-method data analysis. For mixed-method analysis, the teacher qualitative responses from Phase II were combined in an Excel spreadsheet with the quantitative question counts and the demographical information by gender, length of teaching experience, and subject taught. By using the spreadsheet (see Appendix H) and uploading it into the R software, there was an opportunity for a much more comprehensive statistical analysis for Phase II.

In Phase II, this study integrated quantitative data in an interactive way with the qualitative insights, allowing for multiple layers of analysis. “Increasingly, mixed methods approaches have gained favor as an alternative to reliance on either a qualitative or quantitative orientation. . . to provide a comprehensive methodology for the research purposes” (Kington, Sammons, Day, & Regan, 2011, p. 106). Figure 11 illustrates the basic steps taken in this mixed-

method analysis for Phase II. In Phase II, both a quantitative and qualitative Excel spreadsheet (see Appendix H) data was uploaded into the R program. R facilitated data display processes, charts, tables, and diagrams. The data analysis using R was supported by R tutorials and consultation with the BYU Statistics department.

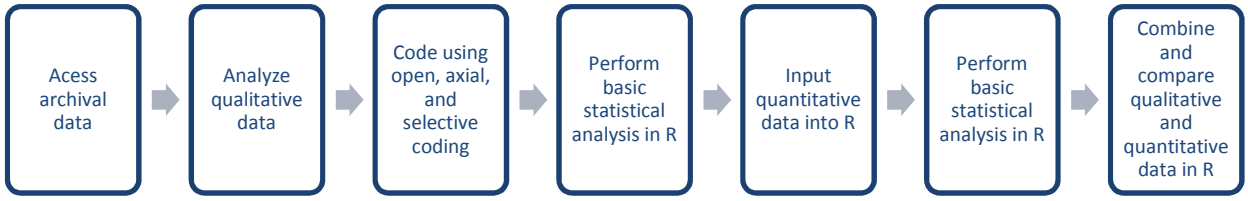


Figure 11. Steps in data analysis for Phase II.

The coded qualitative data could now be explored for relationships with participants' quantitative attribute profile (gender, level of experience, subject taught). Combining qualitative and quantitative data, was beneficial in facilitating the synthesis of the mixed methods data from multiple sources, adding another dimension to the data, and identifying patterns across the methods (Andrew, Salamonson, & Halcomb, 2008).

Classification trees and regression analysis. One simple, yet powerful analytical tool used to determine the most important variables in a particular dataset is classification trees and regression analysis (Morgan, 2014). R software, the most widely used software for this predictive statistical method (Morgan, 2014), classification tree analysis and linear regression models were used in this study to analyze the combination of quantitative and qualitative data. "Classification tree analysis is a potentially powerful tool for investigating multilevel interactions" (Camp & Slattery, 2002, p. 813). Classification trees showed how the response variable or target variable from Question 1 (teacher value) could be predicted by explanatory variables (all other questions) (Morgan, 2014). Classification trees can cope with a mixture of

variable types while not "requiring stringent theoretical or distributional assumptions of the more traditional methods" (Camp & Slattery, 2002, p. 814). Classification trees screened for the predictor variables from the other question responses that are the most important in relation to teacher value (Camp & Slattery, 2002). Or in other words, the classification trees suggested which questions were correlated with how valuable the teacher found student evaluations.

One limitation with classification trees is that they *overfit,* which means that they tend to assume that all the data will behave in a certain or a similar way for all teachers. Since this is not a reality, the R software uses a statistical technique called "bagging" that creates multiple similar data sets (Morgan, 2014) using some variables of interest. The algorithm creates a classification tree based on those randomly chosen variables and the algorithm uses a portion of the data that it has set aside to see how well it has fit the data. Those randomly chosen variables are identified as either important or non-important. This process is repeated many times until there is a statistical forest of random trees. A random forest loses interpretability, but it does offer insight into which variables are important, related, and which ones are not (Santorico & Shan, 2010).

Multiple linear regression was used to predict the importance of the target variable (teacher value) in relation to the importance of two or more other variables for the purpose of discovering possibilities, theory building or idea generating rather than verifying or testing (Morgan, 2014).

Member checking and saturation point. The credibility of the findings, based on the selective coding, was verified through a collaborative process of member checking. After the data was coded and the themes outlined, together with the quantitative analysis, the researcher presented again to the faculty as a whole, seeking participant validation. This "active interaction between the inquiry and the research participants" (Cho & Trent, 2006, p. 320) acted as a form of

member checking. Member checking "occurs throughout the inquiry, and is a process in which collected data is 'played back' to the informant to check for perceived accuracy and reactions" (p. 322), a process of taking the study back to the participants and asking them about the accuracy of the report (Creswell, 2002, p. 280). Phase I general results and Phase II results were presented to the faculty at the end of each school year for verification.

As there were differing opinions among the faculty as to the value of the student input and as to which issues were of the highest concern, Lincoln and Guba (1985) believe participants "may be able to agree that reconstructions are fair even if they are not in total agreement with them" (p. 315). The relationship of trust between the faculty and researcher and within the faculty was already established and was a benefit in ensuring respect for differing opinions and ease in revisiting findings as well as encouraging, and verifying that all viewpoints had been represented. The whole faculty discussion created a sense of agreement on the findings of overall teacher value in student evaluations while still also agreeing that concerns existed on how student evaluations will be used.

After revisiting and rechecking of the data, the researcher implemented an additional Phase III because the data revealed that further investigation was needed. A saturation point had not been reached. When the results of these additional insights from Phase III were added, the researcher determined a saturation point "when no more new information surfaces, evidence of themes being fully developed" (Creswell, 2006, p. 273) for this particular study and focus had been reached.

Chapter 4

Results

This chapter details and displays the findings or results uncovered from the qualitative and quantitative data collected from Phase I, Phase II, and Phase III from one high school in Utah according to the mixed-method research design described in the methods section. The target population was a small, purposeful sampling, coming from one high school with a high participation rate within that context.

Phase I

Quantitative data for Phase I. The data from the four quantitative questions on teacher questionnaires was manually entered into the R program software (see Appendix G) because in this phase the initial teacher survey was completed by paper and pencil hard copies. A bar graph illustrating the basic frequency of teacher responses for each question is shown in Figures 12–15 and Table 3.

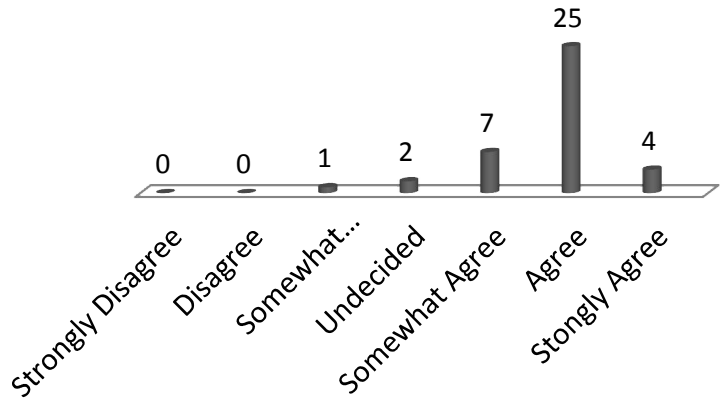


Figure 12. Frequency of response question 1: Overall, I found the student survey experience valuable.

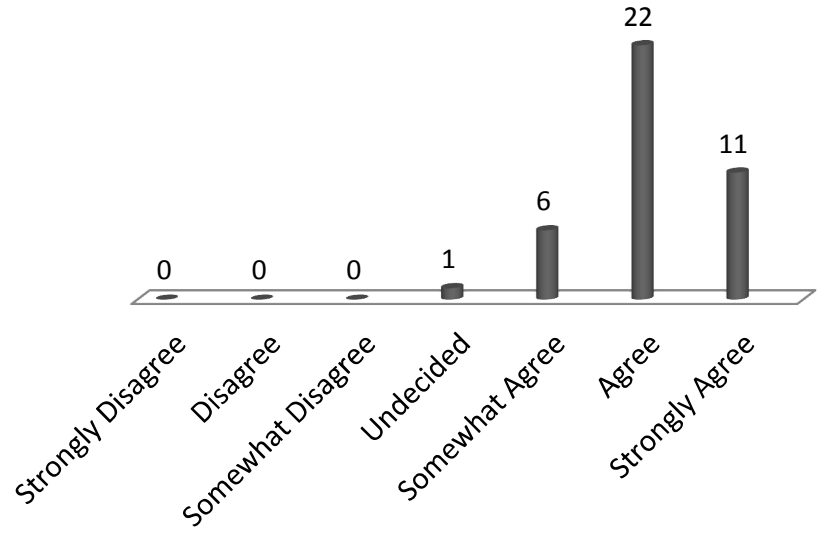


Figure 13. Frequency of response question 2: I will make changes in my instruction due to student input from the survey.

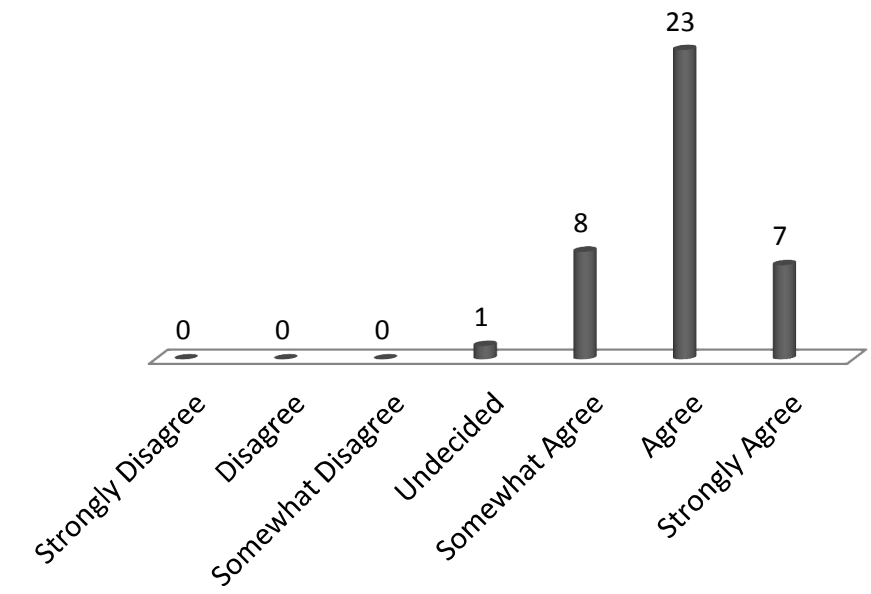


Figure 14. Frequency of response question 3: The student survey facilitated reflection on my practice.

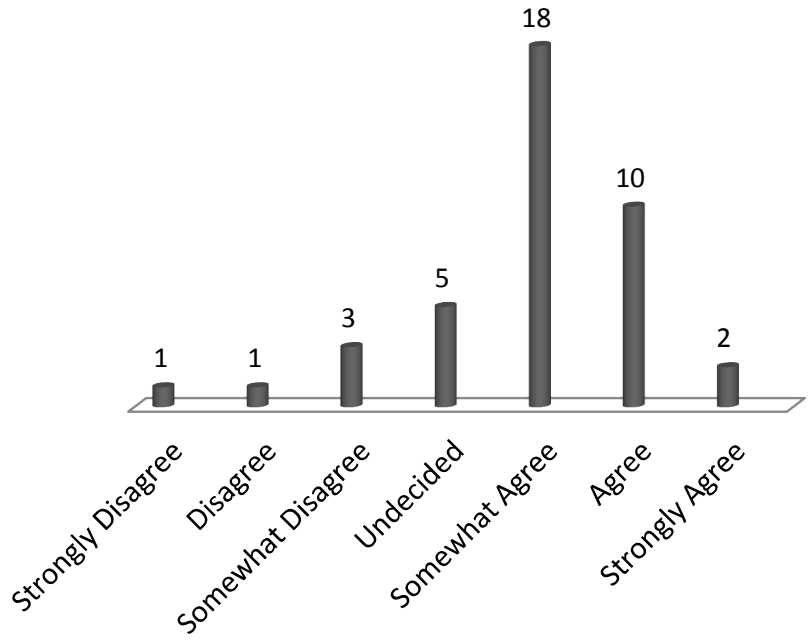


Figure 15. Frequency of response question 4: I believe students are qualified to rate my instruction.

Using the frequency data, Table 3 displays and compares the mean, variance and response distribution for each question. Variance is reported rather than standard deviation because standard deviation is usually used when the sample is being used as representative of the whole population (all teachers). In this case study of a particular faculty, the results may or may not be representative.

Table 3

Mean, Variance and Response Percentage Distribution for Four Questions 1-4 Phase I

Question	Mean	Variance	1	2	3	4	5	6	7
1	5.67	0.79	0	0	2.5	10	15	62.5	10
2	5.80	0.63	0	0	0	5	27.5	50	17.5
3	5.95	0.56	0	0	0	2.5	22.5	52.5	22.5
4	4.67	1.71	2.5	5	12.5	10	45	22.5	2.5

Over 87% of the teachers indicated on Question 1 that they found some value in the student evaluations with 73% in either agreement or strong agreement. The mean for Question 1 was 5.7 out of 7 or in the high range of finding some value. Only one teacher indicated slight disagreement with the value and two were undecided, meaning the remaining 37 found some value in the student perspective.

Even more favorable was the response for Question 2, or indication that the teacher would be changing instruction based on the student input with 95% of teachers were in some level of agreement. Five teachers were undecided and not one teacher was in disagreement. With a mean of 5.8, overall teachers indicated a more positive response to changing instruction than finding value.

Results for Question 3, the level of reflection facilitated by the student input, received the most favorable response. Over 97% of teachers indicated the student responses caused at least some reflection on their teaching. The mean for Question 3 was a 6 out of 7, or in the range of full or strong agreement. In fact, not one teacher disagreed and only one was undecided on the power of the student surveys in causing them to reflect. Teachers responded higher on having reflected upon the evaluations than finding value in the evaluations.

Question 4 had the lowest overall mean at 4.7, below the agreement level. Thus, more teachers disagreed that students are qualified to rate their instruction. Still 70% of teachers overall showed some agreement, but only 2 teachers were in strong agreement compared to 4 for Question 2, 7 for Question 1, and 11 for Question 3.

Figure 16 displays the level of strong agreement for each question. Showing that even though Questions #1, 2, and 3 had means within the agreement level, there is still a difference in the level or number of strong agreement responses. Question 3, teachers who reflected on practice due to the student surveys, had the highest level of strong agreement. Question 4 had the lowest level of strong agreement.

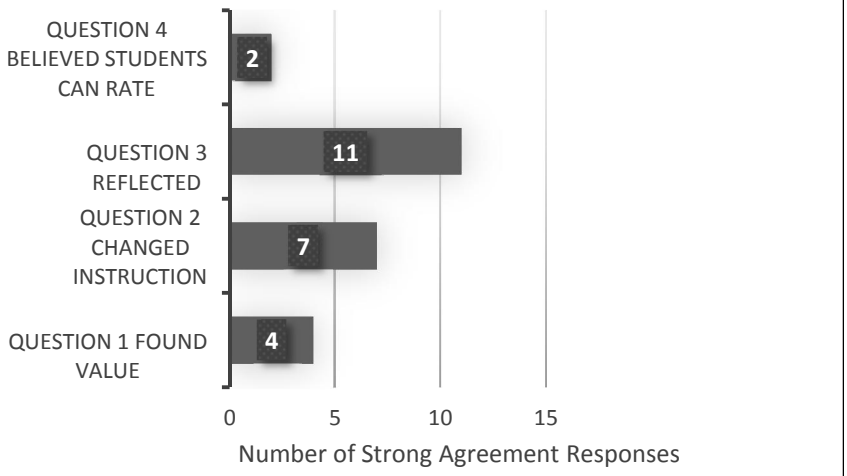


Figure 16. Level of strong agreement by question Phase I.

Figure 17 compares the variance by question. Question 4 showed the greatest degree of variance. The data in this question was spread out across all possible responses and had a lower concentration in the positive agreement responses, showing that there was overall a greater degree of disagreement and a less positive response on this question.

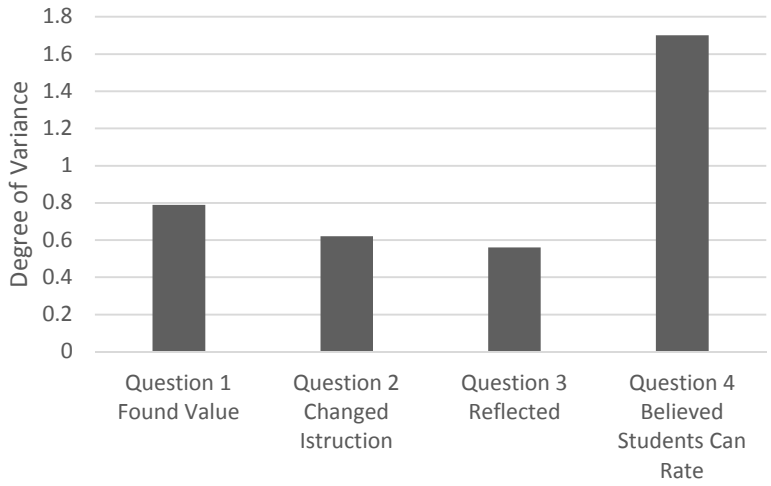


Figure 17. Degree of variance by question Phase I.

For Phase I, the use of classification trees and the bagging method showed that all of the variables were statistically important in relation to the main research question of teachers finding value in student surveys. The explanatory variables or Questions #2, #3, and #4 were important if the score was greater than zero and the more positive the number, the more important. Thus, based on the data shown in Table 4, questionnaire Questions #2 (Changed Instruction), #3 (Reflected), and #4 (Believed Students Can Rate) are all related to Question 1 (Found Value), the value level teachers found from student evaluations. Specifically, Question 4 was the most important with the highest score or the stronger relationship to value. This again points to a strong correlation between a teacher's value of student evaluation and a teacher's belief in the ability of students to evaluate or perhaps in a teacher's concern over how that student rating will be used or interpreted.

Table 4

Importance Scores in Relation to Teacher Value (Question 1) Phase I

Question	Importance score
Question 2: Change instruction	28.3
Question 3: Reflected	41.5
Question 4: Students qualified to rate	59.2

A multiple linear regression analysis was used with Question 2 (*Changed Instruction*), 3 (*Caused Reflection*), and 4(*Students Qualified to Rate*), since all were found to be important, as the predictor variables to determine if any of these variables would be *significantly* important as related to Question 1 or teacher value. Regression coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant. This statistical control that regression provides is important because it isolates the role of one variable from all of the others in the model (Morgan, 2014). To interpret these results consider that for a one unit increase in Question 2 (*Changed Instruction*), the response to Question 1 (*Found Value*) should increase by x amount. For example, the coefficient for Question 4 was 0.26. So in context, for a one unit increase in Question 4 (going from agree to strongly agree), the likelihood of a teacher finding the student surveys valuable increases by .26. Or in other words there was a positive relationship. Those teachers that responded more positively to Question 4 responded more positively to Question 1. Table 5 presents the coefficients for each variable, along with the confidence intervals for these coefficients from this regression analysis. If these confidence intervals do not contain zero, then they are statistically significant, meaning that if this experiment been repeated many times, then the interval shown would contain the true coefficient 95% of the time.

Table 5

Regression Coefficients for Multiple Linear Regression as Related to Teacher Value Phase I

		Confidence intervals	
Question	Coefficient	2.50%	97.50%
Q2 Changed Instruction	0.30	-0.03	0.64
Q3 Reflected	0.30	-0.04	0.65
Q4 Believed Students Qualified to Rate	0.26	0.08	0.44

Thus, Questions #2 (Changed Instruction) and #3 (*Caused Reflection)* were important, but not statistically important because of the lack of a sufficient confidence interval. However, the coefficient for Question 4 (*Students Qualified to Rate)* was a positive 0.26 with a 95% confidence level, suggesting that there is a statistically significant positive relationship. Those teachers that responded more positively to Question 4 were more statistically likely to respond more positively to Question 1. Teachers' value for the student input was statistically related to teachers giving credence to student ability to offer input.

Qualitative data Phase I. The initial analysis for Phase I qualitative data occurred before Phase II so that the second phase would represent the learning that had been facilitated by the first phase. These changes and revisions are described in the archival data description of the case study. For the purpose of the research study, the researcher conducted some additional analysis for Phase I. From the coding process described in the Data Analysis section, Table 6 was created using the identified core or central themes and key responses from questions that had not been previously analyzed. All teacher comments were included or placed in a category with no response marked as none. Theme creation collapsed and combined the categories, but did not exclude any comment. Some teachers gave multiple responses to one answer and these were counted in each category. Teachers left some questions blank, which was the only limiting

factor. Even with some questions left blank, the comments included were those that the teachers felt strongly about, either in the negative or positive, to take the time to write down. The analysis then may weigh heavily on feelings of both extreme sides of the arguments while failing to accurately represent those in the middle, a phenomenon not uncommon in political arenas. However, all teacher participants responded or *voted* in the quantitative responses and thus, again, the balance of a mixed-method analysis.

Table 6

Central Coded Themes from Qualitative Questions for Phase I

Question	Coded theme	Response %	Key response
1. What insights did you gain from the data? 2. Did you gain any valuable information from student comments section?	Facilitated Reflection	5%	"I don't think I have ever done anything that has caused more reflection including looking at test scores."
	Gave Feedback on Specific Assessments, Methods, or Projects	7%	"I'm not going to do _____ projects again (really, I already knew this, but it confirmed what I thought)."
	Gave Feedback on Classroom Procedures or Management	12%	"I do a monthly calendar for my advanced classes and I know they really appreciated it. I have never done that for my regular classes - an ah ha moment. Next year, I will do a monthly calendar in all my classes"
	Highlighted Need for Increased Individual Attention or Additional Help for Struggling Students	10%	"I was surprised at how many students felt like they did not receive help when they needed it. I need to make sure I am clear on what to do if they need help."
	Increased Understanding of Student Awareness	5%	"Students are watching all of the time."

Question	Coded theme	Response %	Key response
4. Will you make any changes in instruction because of the results of your student evaluations? What specific change will you make?	Add daily review	7%	
	Improve classroom management (decrease noise, downtime, etc.)	20%	
	Add enrichment	7%	
	Give more quizzes (formative assessments)	2%	
	Decrease visible teacher frustration	12%	
	Update grades and communicate student progress more frequently	30%	
	Make objectives clear	5%	
	Take time to build relationships	16%	
	Use more examples	1%	
	Implement more hands-on activities	2%	
	Beware of favoritism	2%	
What are your biggest concerns with student evaluations?	Contradictory/range comments	3%	
	Hurtful comments	10%	
	Teacher personality judgment	3%	
	Core vs. electives	2%	
	Content complaints/can't change core	2%	
	Subjective, not objective	5%	
	Time of day/year	15%	
	Teachers blamed for student apathy	25%	
	Lack of student knowledge	2%	
	How will results be use? Punitive?	42%	

Phase II

Phase II quantitative data. The data from the six quantitative questions for Phase II was entered into an Excel spreadsheet and then uploaded into the R program software. Similar to the data display for Phase I, the following figures provide bar graphs illustrating the basic frequency of responses for each question with an additional table that displays the means, frequency counts, and variance for each Phase II Likert response question.

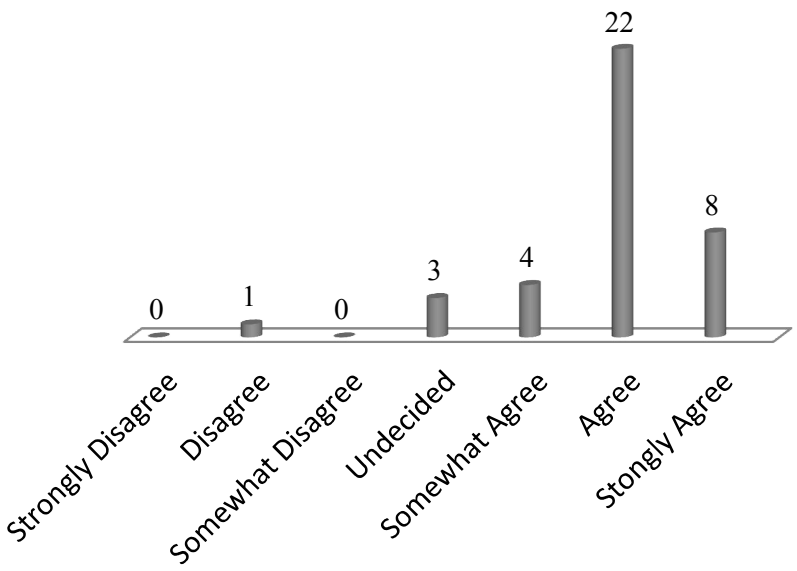


Figure 18. Frequency of response question 1: Overall, I believe student evaluations are of value.

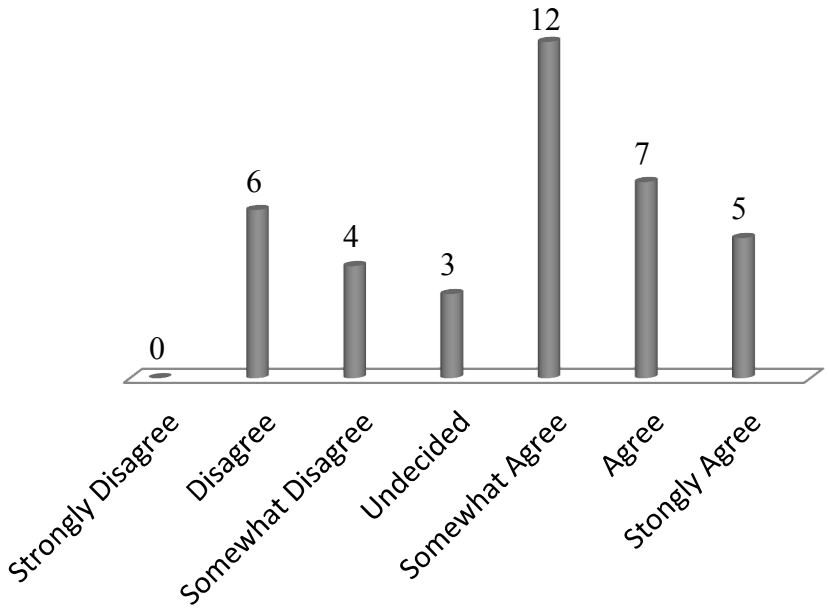


Figure 19. Frequency of response question 2: I made changes in my instruction due to student input from last year's student evaluations.

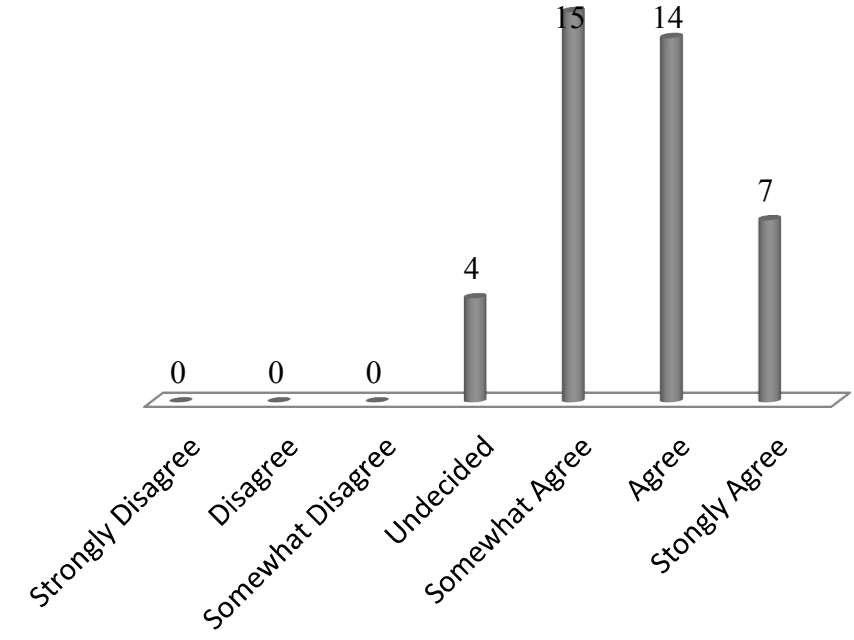
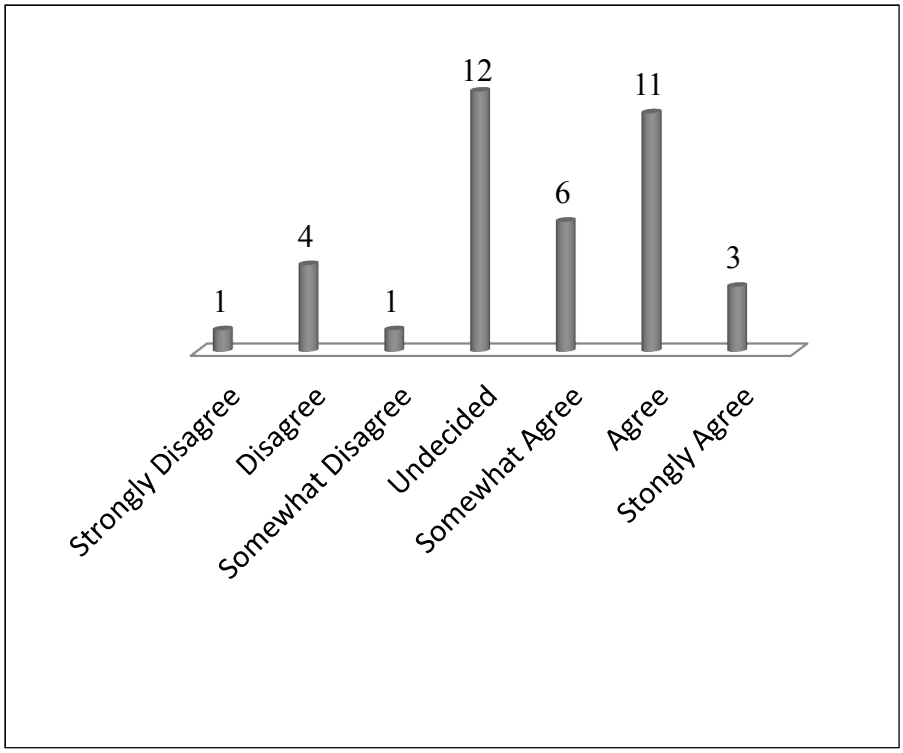


Figure 20. Frequency of response Question 3: I used input from student evaluations last year to facilitate classroom discussions this year.



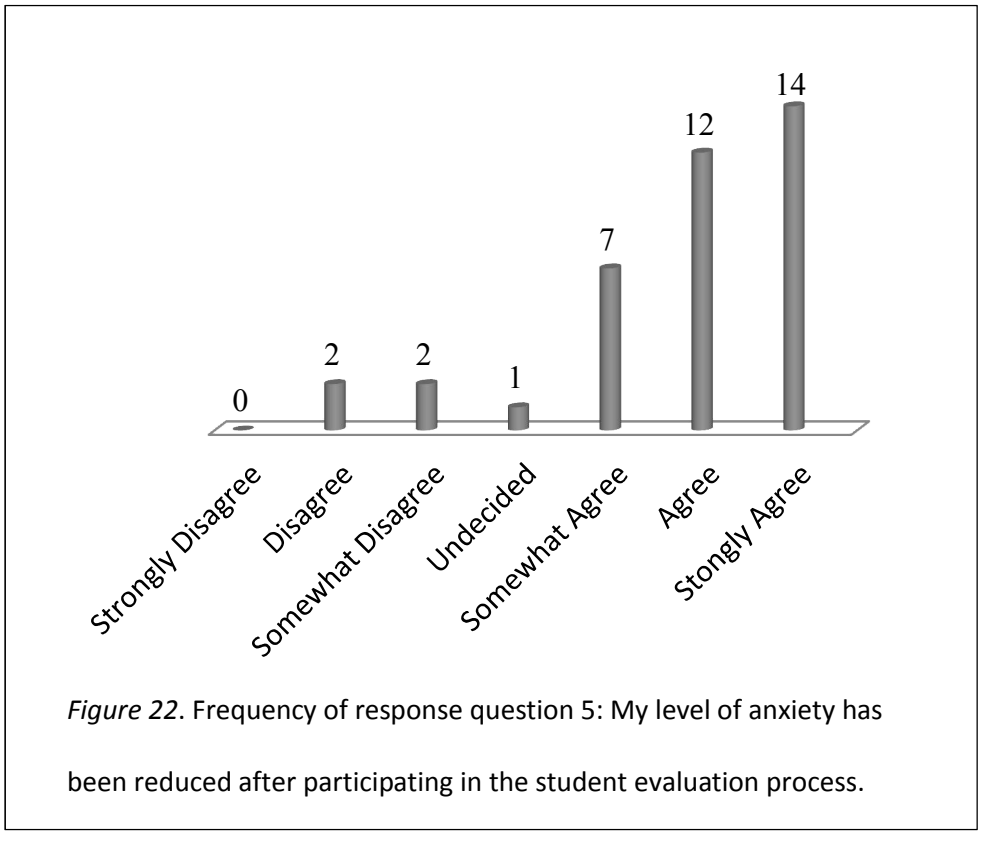


Figure 22. Frequency of response question 5: My level of anxiety has been reduced after participating in the student evaluation process.

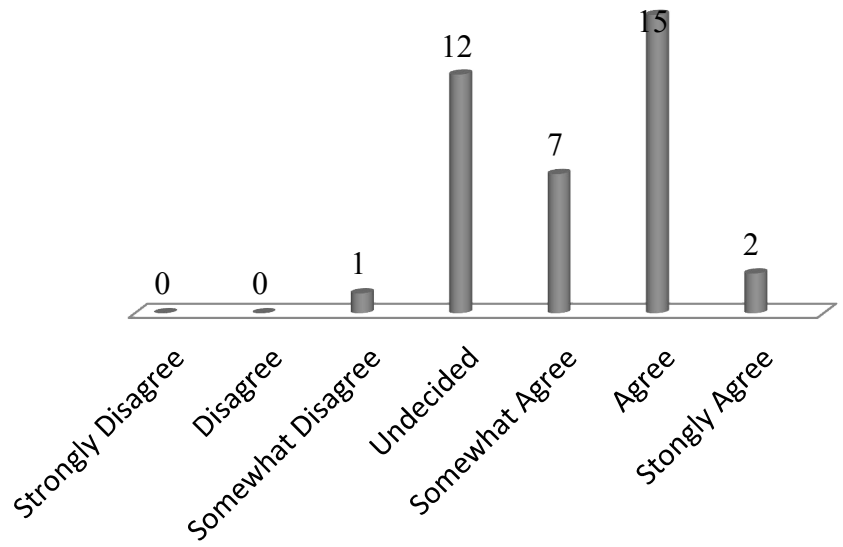


Figure 23. Frequency of response question 6: The collaborative nature allowed in the implementation of student evaluations increased my level of support for the process.

Table 7

Mean, Variance and Percentage for Quantitative Questions Phase II

Question	Mean	Variance	1	2	3	4	5	6	7
1	5.8	1.0	0.0	2.7	0.0	8.1	10.8	59.5	18.9
2	5.7	0.7	0.0	0.0	0.0	5.4	40.5	35.1	18.9
3	4.6	3.0	2.7	16.2	10.8	8.1	29.7	18.9	13.5
4	5.1	1.2	0.0	0.0	2.7	8.1	16.2	27.0	45.9
5	4.7	2.2	2.7	8.1	2.7	35.1	16.2	27.0	8.1
6	5.1	1.2	0.0	2.7	0.0	32.4	21.6	37.8	5.4

Question 1 was the same for both Phase I and Phase II, "Overall I believe student evaluations are of value." In Phase II, the level of value identified by teachers increased from a 5.6 to a 5.8 mean score. However, there was one disagree in Phase II and there had been none in Phase I. Perhaps, this came from one of the two new teachers that joined the faculty this year.

Question 2 asked teachers if they had made changes in instruction bases on input from Phase I. Teachers responded favorably with an overall mean of 5.8 which is almost identical to the 5.8 in Phase I for the question asking teachers if they planned on making changes.

For Question 3, teachers indicated whether or not they had had classroom discussions over the results from the student evaluations from last year. This response had the lowest mean for Phase I with a 4.6 mean and the greatest degree of variance or responses.

Question 4 was the same as Question 3 in Phase I where teachers indicated their level of reflection over student input. In Phase II, the level of agreement went down to a mean of 5.1 compared to a mean of 6 in Phase I. It makes sense that the initial phase or the first time might produce stronger reflection.

Phase II included the additions of Questions #5 and #6. Question 5 asked teachers to indicate whether they had experienced a decrease in their anxiety level for Phase II as compared to anxiety levels they felt during the initial student evaluation process of Phase I. Teachers responded with an overall mean of 4.7. The number of teachers with an "undecided" response was the highest for this question. There is evidence that anxiety was reduced. A better wording of the question might have been to ask whether anxiety had decrease or increased with an open-ended question asking for an explanation.

In Question 6, teachers were asked to indicate whether the collaborative nature of the implementation of student evaluations made them feel more supportive of the process. With a mean of 5, the teacher response was overall positive, but 32% were still "undecided." Additional insights were found with the results of quantitative Question 6 on the collaborative effort. In the department focus groups, when this question was discussed. The teachers asked for a clarification of what was meant by a "collaborative process." The discussion comments indicated

that teachers did find value in the collaborative process or efforts when clarified by the researcher as implementing the mandated evaluations as a cooperative venture. Teachers did not see the legislative mandate as a collaborative process which they clarified was indicated by some of their responses. When asked to elaborate, teachers stated that they found value in collaborating over the results, liked having a say in the type of questions asked the students, and liked knowing the questions that would be asked ahead of time.

Phase II qualitative data. The teachers in Phase II completed a list of 10 new open-ended questions. These questions asked teachers to make comparisons between the two phases. The questions were designed to help discover differences and similarities between an initial year and a second year of going through the student evaluation process (see Table 8). Coding for Phase II went through the same process as Phase I as described in the Data Analysis section, but every question was analyzed because no previous analysis had been done. Given the small population and the ability to identify participants, additional demographic information was not included in the key responses to protect the promised anonymity.

Table 8

Phase II Teacher Qualitative Analysis: Coded Themes

Question	Coded themes	Response %	Key response
1. How did your evaluation results differ from last year?	More positive	54%	"Comments better this year."
	The same or none	41%	
	More negative	5%	
2. Identify changes you noticed due to question revisions from initial evaluation.	Process quicker	5%	
	More positive	27%	
3. What additional changes will you make in instruction because of the results of this year's student evaluation?	Let students know I care	19%	"Increased personal interaction."
			"I recorded lectures for absent students."
	Adjusted workload	11%	"I reworked many assignments."
	Increased communication	3%	"I recorded lectures for absent students."
			"I gave more short term quizzes."
	Increased student driven instruction	5%	"I tried to appeal more to student interest."
			"I identified areas where students feel like they need more instruction."
			"I gave student survey of my own after 1st semester and implemented changes 2nd semester on my own."
	Increased variety of teaching strategies	5%	"I tried to be more exciting."
			"I will try not to get into a rut and keep doing the same thing."
	Increased preparation	2%	
	None		
	Increase classroom management	16%	
	Increase hands-on activities	2%	"I implemented more activities rather than me giving notes all the time."
	Make timely updates to grading programs	36%	
	Make self-improvement changes (patience, frustration, etc.)	5%	"I will be nicer and start each day fresh without bringing in any negative feelings about students from the day before."
			"Watch my frustration level."
	Make sure all students receive help	8%	

Question	Coded themes	Response %	Key response
	Continue changes from last year	16%	
	Classroom management	8%	
	Add variety	5%	
4. What did you find the most valuable from student evaluations (either year)?	Feedback	42%	"Students who may not always speak up get a chance to voice an opinion."
	Validation	19%	"This year it was nice to know that students noticed the areas that I had made an effort to change."
	Reflections	11%	"Got me out of my comfort zone."
			"Student comment section by far the most valuable."
5. What did you find the least valuable from student evaluations (either year)?	Irrelevant comments	11%	
	Student immaturity or inability	32%	"Student comments were too vague. I want them to be specific on how I can improve."
	Conflicting responses	8%	
6. What is still your biggest concern?	Punitive	22%	"Feel like these will be used to determine what type of teacher I am."
			"Our subject very different than teaching a chosen elective."
	Other	11%	"Too many what if's."
	Validity	5%	
7. What did you notice by not having as many classes give feedback?	Not representative	46%	"Did not get as much feedback with fewer classes surveyed"
	Fewer comments to read	5%	"Hope final state decision includes feedback from all classes."
			"Not enough feedback from all of my classes so that I can make real positive changes."
8. What ideas do you have on how to make additional changes to the implementation process?	Other	11%	"Make survey time systematic-on a set schedule."
	Earlier	26%	"Needs to be earlier in the year-before testing. Students are burned out."
9. Any other information or comments?	(Comments added to other themes)		

In Question 1, 54% of teachers indicated more positive student evaluations in Phase II as compared to Phase I. Perhaps, this change is due to the changes that 90% of teachers indicated in quantitative Question 2. Some teachers, 27% , however, indicated that more positive student evaluations were due to the question revisions from Phase I to Phase II.

When looking at the responses to Question 3, referring to the specific changes the teacher made based on last year's evaluations compared to teacher responses to the changes that teachers indicated they would make based on open-ended Phase I, Question 5, the lists are very similar, which indicates that teachers both planned and followed through with changes in instruction. Noting, however, that some teachers planned on adding enrichment and making objectives clearer in Phase I, there were no teachers who reported making these specific changes in Phase II.

In Phase I, 95% of teachers indicated they would make changes based on student feedback (Question 2) and in Phase II, 90% of teachers reported actually making changes (Question 2). The comparison of this data, changes predicted and changes made (see Table 9), showed that teachers had not only responded to student input, but also followed through with intended changes or at least indicated that they had made changes.

Table 9

Correlation of Changes Predicted in Phase I and Changes Made in Phase II

Phase I—changes predicted	Phase II—changes made
Improve classroom management	Classroom management
Take time to build relationships	Letting students know I care
Decrease visible teacher frustration	
Beware of favoritism	
Add daily review	More preparation
Give more quizzes	
Implement more hands on activities	More student driven/engagement
Use more examples	Variety of teacher strategies
	Adjust workload
Update grades and communicate student progress more frequently	Better communication/update progress
Add enrichment	--
Make objectives clear	--

At the end of the department focus group session, given that teachers indicated making changes in updating student progress, an insightful conversation arose naturally as the principal noted that the overall lowest score for both implementation phases for teachers school-wide on the student evaluations was on the topic of teachers keeping students up to date on their progress. Teachers wondered if this meant updating SIS (the school grading program) or if students wanted to be told personally or given print outs. The suggestion was made to make the question clearer by asking "Does my teacher update SIS regularly?" while at the same time adding the question, "Do I check SIS regularly?" to the student self-assessment questions. The principal also asked teachers to discuss this student concern in PLC's and give him input for a possible school-wide policy.

The detailed notes taken by the principal from the department head focus group offered additional clarifications for some of the qualitative questions (see Table 10). The comments came from or were created by the discussion and interaction between the participants which demonstrated the ability of focus groups to generate new insights.

Table 10

Qualitative Data from Department Head Focus Group Phase II

Clarifying comment	Corresponding question
"What do you mean by collaboration?"	Quantitative question 6
The collaborative effort did help in the interpreting the data	
"I liked having had a say in the questions."	
"Knowing the questions in advance was a good thing."	
"Perhaps less informative this year than last year. Fewer classes surveyed."	Qualitative question 1
"Yes, the comments this year were better than last year. Changes made a difference."	
"Seeing the student take notice of things is valuable."	Qualitative question 1
"I saw a difference in my beginning computer required classes and upper level computer electives."	Qualitative question 7
"Most teachers do get an occasional stinger. Worried on where they will go."	
"A variety of classes bring a variety of responses- better to survey more classes."	Qualitative question 7
"That student comment was actually a negative, but I took hard class as a positive."	
"If you discipline a kid often, they will usually give you bad results."	
"A bell curve might be a good way to look at it."	
. "In our PLC meetings, several of us talked about asking students questions about the evaluations."	Quantitative question 3 and qualitative question 3
"Class size was a variable that needs to be looked at perhaps."	Qualitative question 6
There was a general consensus that the teachers would have liked to have had all of their classes surveyed and earlier in the year.	Qualitative question 8
"It's hard to have a one on one conversation when you have 40 kids in a class."	Qualitative question 3
The question on "Does my teacher keep me updated on my progress?" Suggestion to change it to "Does my teacher keep SIS up to date?"	Qualitative question 9
"Add to student reflection. How often do you check SIS?"	
"Appreciated administration censoring crude comments."	

The idea of looking at the results of student evaluations falling within a typical bell curve was a new thought in this second phase. The comment was shared as a method for de-stressing over evaluations by suggesting some inevitability in student evaluations giving most teachers a "C' grade with some receiving higher and some lower. Given the high level of anxiety teachers demonstrated, teachers felt that this perspective would only lower stress levels if the teacher fell to the right of the curve. Teachers in this group did not respond favorably to the prospect of getting a "C" grade. Teachers still indicated concerns over student unreliability and over possible punitive measures, similar to Phase I. These responses add additional insight into perhaps why teacher anxiety level from quantitative Question 5 did not show a decrease.

Mixed-method analysis. For Phase II, the use of classification trees and bagging was once again used to determine if any correlations would be statistically significant between the quantitative data, the qualitative data and the level of value found in teacher evaluations. Phase II used a combination Excel spreadsheet of quantitative data and coded qualitative data as well as demographic data, thus increasing the opportunity for mixed-level analysis.

When using qualitative variables, all coefficients are measured relative to the base level of one response for that question. For example, in the second year data, the three levels of subject were STEM, FA/Humanities, Other (PE, etc.). In the output, no coefficient exists for FA/Humanities because that is the baseline coefficient.

However, after the first trial run of classification trees, given the large number of variables, which now included all of the qualitative codes and demographic information, and the small population size, it was determined that showing statistical relevance would be difficult. So the importance scores from the first trial run were used to select which variables could be removed and a second classification tree analysis was conducted using the reduced set of

variables. The second run of the data was much more effective in identifying variables of importance. The Tables below displays the second run and variables of importance for Phase II.

Table 11

Importance Scores in Relation to Value (Question 1) Phase II

Question	Importance score
quant2 changed instruction	3.3
quant3 class discussion	-1.8
quant4 reflection	6.6
quant5 reduced anxiety	12.9
quant6 collaborative support	13.3
quant7/experience	-5.0
quant8/gender	-1.7
quant/subject	1.4
qual1differ from last year	-0.5
qual2 question revision	-4.2
qual3 changes made	11.0
qual4 future changes	-8.0
qual5 most value	0.3
qual6 least value	1.5
qual7 concern	1.5
qual8 fewer classes	1.4
qual9 ideas for change	1.4

In Phase II, responses given to quantitative Questions 2 (*Changed Instruction*), 4 (*Caused Reflection*), 5 (*Reduced Anxiety*), 6 (*Collaboration Created Support*) were important in relation to the level of value the teacher placed on student evaluations with positive scores greater than zero. In addition to this, qualitative Questions 3 (*Specific Changes Made*), 5 (*Most Valuable*), 6 (*Least Valuable*), 7 (*Biggest Concern*), 8 (*Fewer Classes Surveyed*), and 9 (*Ideas for Change*) were also important and demographically, subject (subject taught by the teacher) was important. Qualitative Questions 1 (*Differences Between Phases*), 2 (*Changes Noticed due to Question Revision*), 4 (*Future Changes in Instruction*), and demographically, gender, and years of

experience were not significant in Phase II in determining the level of value a teacher placed on student evaluations.

The highest importance scores as they related to the value a teacher placed on student evaluations came from teacher responses to the level of reflection, reduced anxiety, level of support gained from the collaborative process, and the changes made in instruction due to last year's student evaluations.

Once questions were identified as important, a multiple linear regression model was used to analyze the identified important quantitative and qualitative variables together to determine if the relationships between any of the variables found important were statistically significant. Table 12 presents the coefficients for each variable, along with the confidence intervals for these coefficients from this regression analysis. If these confidence intervals do not contain zero, then they are statistically significant, meaning that if this experiment had been repeated many times, then the interval shown would contain the true coefficient 95% of the time.

Table 12

Multiple Linear Regression Importance Coefficients and Confidence Intervals Phase II

Column1	Coefficient	2.50%	97.50%
4 caused reflection	0.5	0.1	0.9
Quant 5 reduced anxiety	0.1	-0.3	0.2
Quant 6 collaborative support	0.0	-0.5	0.4
Subject- other	-0.2	-1.1	0.8
Subject-STEM	0.4	-1.4	0.5
Qual3 communication	0.2	-1.7	1.3
Qual3 students know I cared	0.1	-1.3	1.1
Qual3 none	0.2	-1.0	1.3
Qual3 preparation	0.4	-2.2	1.4
Qual3 student driven	0.3	-1.1	1.7
Qual3 variety	0.4	-1.3	2.1
Qual3 workload	0.7	-0.7	2.1

Using multiple linear regression quantitative variables can be interpreted similarly to Phase I. However, in Phase II there are qualitative variables as well. The coded question choices for each qualitative question become a variable. Thus, the levels, or responses, become the focus of looking for importance. With qualitative variables in linear regression, all coefficients are measured relative to some base level of that question. For example, in the Phase II data, the three levels of subject were STEM (Science, Technology, Engineering, and Math), Fine Arts/Humanities, Other (PE, etc.). In the output, that there is no coefficient for FA/Humanities because it was the baseline coefficient. The coefficient for STEM can be interpreted as follows. Since the coefficient is -0.475 or .5, teachers that teach STEM will on average mark about half a unit lower than teachers in the FA/Humanities section on finding value in student evaluations. There was no importance difference noticed between Other (PE, etc.) and FA/Humanities teachers.

Interpreting the quantitative variables with multiple linear regression means looking for statistical significance. If the confidence intervals do not contain zero, then they are statistically significant, meaning that if this experiment been repeated many times, then the interval shown would contain the true coefficient 95% of the time. As shown in Table, with a positive coefficient of .5 and a confidence interval of 95%, quantitative Question 4 (*Caused Reflection*) shows a statistically positive relationship. In context, for a one unit increase in Question 4 (*Caused Reflection*) (going from agree to strongly agree), the likelihood of a teacher finding the student evaluations valuable, or Question 1 (*Teacher Value*), also increased. At least for this population, those teachers that indicated that evaluations caused reflection were more likely to respond that they found value in student evaluations. Teachers who valued the student input are statistically related to teachers reflecting on student evaluations.

Phase III

Quantitative data analysis Phase III. A survey on Qualtrics software was emailed out to the faculty. The survey contained a copy of the previous question from Phase I #4 *(Students are Qualified to Rate Teacher Instruction*) and the new question Phase III (Students are Qualified to Give Feedback), only asking teachers to note the wording change from "rate" to "give feedback" and indicate their agreement level on the new question. The responses were anonymous. The analysis would compare the teacher response to the wording change, looking to see if there was a significant difference between the teachers' response to the new question. A frequency response chart for Phase III was created and compared to a frequency of response for Phase I with the original question (see Figure 18). A table comparing the means and variance for each question was also created (see Table 13).

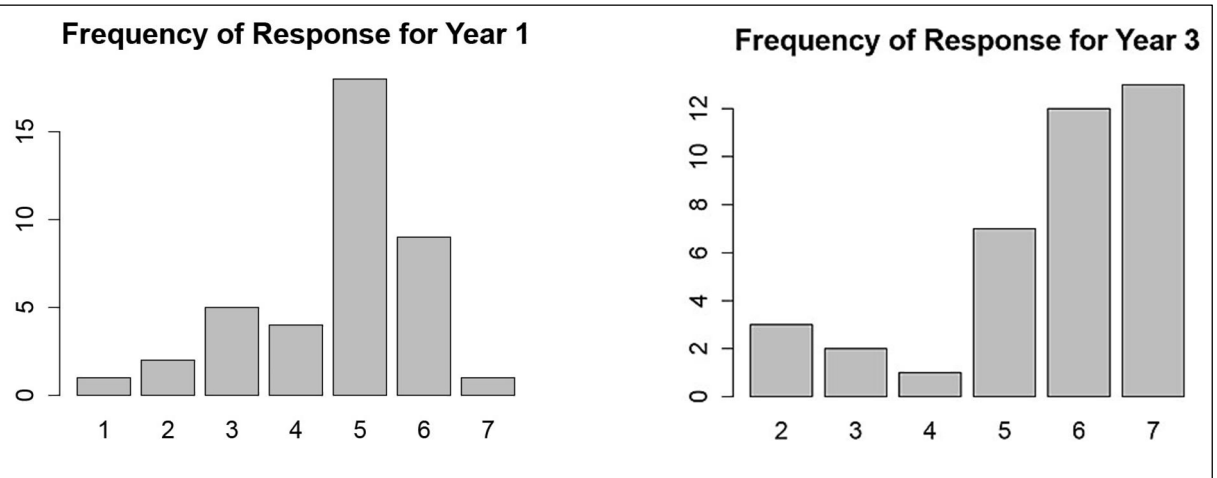


Figure 24. Comparison of frequency response between Phase I and Phase III.

Table 13

Comparison of Mean and Variance for Question from Phase I and Phase III

Phase	Mean	Variance	Standard deviation
Phase I	4.7	1.7	1.3
Phase III	5.6	2.3	1.5

Even with very little explanation and just re-asking the question, the results initially showed an increase in teacher acceptance of this new wording. The results still showed variance, even more for Phase III. The results indicated an increase in the average mean by a whole point from Phase I to Phase III. Further testing had to be done to see if the findings were statistically significant.

The researcher used a two-sample t-test to see whether there was a statistically significant difference in the responses to the questions from Phase I to Phase III. A second verification was done using the Mann-Whitney U test. Tables 14 and 15 provide the phases comparisons, test statistics and p-values. A low p-value means that there is a low probability that the observed differences were purely by chance. In statistics, the rule of thumb is that a p-value less than .05 is generally considered statistically significant.

Table 14

Two-Sample t-Test

Phase	*t*-statistic	*p*-value
Phase I vs. Phase 3	2.99	.0004

Table 15

Mann-Whitney U *Test*

Year	*u*-statistic	*p*-value
Year 1 vs. Year 3	418.50	.0004

Both tests showed that the difference between Phase I and Phase III was statistically significant. This finding means that teachers were more likely to respond in a higher level of agreement for the question with the word *feedback* than they were for the word *rate*. Teachers were statistically more likely to give a high level of agreement to students being qualified to give feedback on teacher instruction than they were to students being qualified to rate teacher instruction. The change in wording had made a statistical difference in their responses.

Summary

Table 16 summarizes the findings based on all three phases of the research study directly correlated to the stated research purpose and questions. The initial questions of the researcher guided the study, but the insights, if not answers, emerged from the data and the process of the research itself.

Table 16

Conclusions from All Three Phases by Research Question

How do teachers react to SETs?	
Will teachers find value in SETs?	
Value level	Despite anxiety, teachers demonstrated a high level of agreement, even an increase from Phase I to Phase II, in the value of student input on their instruction. However, the level of value was related to the level of anxiety with higher anxiety creating lower teacher value. Thus, lowering teacher anxiety could be the key to increasing teacher value.
Reflection level	Teachers responded favorably to student evaluations causing teacher reflection. Teachers indicated an even higher level of reflection than value.
Actual impact on teacher instruction	Teachers responded to student voice by making changes in instruction based on student input. Anticipated changes in Phase I directly corresponded to the actual changes indicated by teachers in Phase II. Teachers noted in Phase II feeling that more positive student surveys were due to teacher response to initial student surveys.
Anxiety level	Teachers indicated on teacher reflections and in focus groups and departmental discussions of both phases a constant level of anxiety over teacher evaluations. Anxiety was decreased after an initial trial run of the process. Teacher concerns however remained focused on the uncertainty of how student surveys will be used on teacher evaluations and to what extent teacher pay or reputation would be effected. Giving adolescents the power to evaluate instruction caused teachers to question the reliability and validity of student input as an equitable tool for teacher rankings.
What can be done to improve the implementation of SETs?	
Increase value and reflection	The level of teacher value and reflection were shown to be related, increasing either one would increase the other. Phase III uncovered a potential way to increase teacher value/reflection by initiating student input as feedback rather than a rating. Teachers also suggested that timing of the student surveys and coverage level of classes may decrease concerns and therefore increase value.
Reduce anxiety	Because the level of anxiety was found to be related to the level of value and the acceptance of students' ability to give input was also related to teacher value, the finding of Phase III of teachers being more accepting of students' ability to give feedback rather than ratings could be used to design a more effective collection of and use of student input.
Increase impact on instruction	The level of value was related to the degree that teachers made changes in instruction or incorporated student voice. If value was increased by decreasing anxiety, the impact on teacher instruction may also increase.

Chapter 5

Discussion, Conclusions, and Recommendations

After combining the data, interpretation of data was summarized by reporting on key findings, possible explanations for results, and identifiable limitations (Creswell, 2002, p. 252) as the data pertained to the main research questions for each phase. This research on a particular population can inform local practice as implementation methods for teacher evaluations may be left to local control. Under principles of transferability rather than generalizability (Erlandsen, 1993), the findings may also inform practice for similar schools or inform further research for local, state, or national implementation. Despite the small population studied, the results may have the potential to offer insights into the teachers' viewpoint of implementing student evaluations, the triumphs as well as trials of giving students voice.

Discussion

Phase I. Surprisingly, despite the overall high level of teacher anxiety observed by the researcher during Phase I over the student evaluation process and as indicated in focus group and departmental discussions, the results were markedly positive for *Finding Value* in the student evaluation process. Curiously, 12 teachers (30%) indicated a higher level of agreement to *Changing Instruction* than *Finding Value*, possibly suggesting that some teachers had some reservations on value of student evaluations, but were planning on making changes due to the student input. However, this finding could also mean that the continued anxiety level of teachers may keep them from finding value in the process because of concerns over how the results will be used or interpreted by the public. Teachers indicated in focus group discussions and quantitative data that they were willing to and did make changes in instruction based on student evaluations, but were still concerned over the political value that would be put on student

evaluations. These concerns were reported in the open-ended questions and focus group discussions. The qualitative data shows the largest level of concern (40%) came from teacher fear over punitive measures and teachers being blamed for student apathy (25%) becoming associated with student evaluations. If the anxiety level or concerns for punitive measures could be decreased, teachers may find more value in student evaluations.

Also interesting to note, 37% of teachers responded higher on *Reflected on Evaluations* than *Finding Value* in the evaluations, perhaps signifying that causing reflection was not considered the same as finding value. This might be also again be representative of teachers' fears over using evaluations limiting the overall value for some, but not limiting their reflection on student comments.

The lowest teacher response came over the *Ability of Students to Rate* teacher instruction, revealing a central teacher concern over either the ability or the appropriateness of placing students in an evaluative position over teachers. Focus group discussions also centered on this issue with the list of concerns verifying the teachers' anxiety over the validity and reliability of an adolescent's ability to judge or *rate* instruction. This finding may have uncovered a key factor in what strikes at the heart of teacher frustration over student evaluations and creates a possible obstruction foiling the development of a student-teacher partnership in the classroom. Given that *Find Value* for student evaluations was significantly related to *Frustration over Student Evaluative Abilities*, this finding may suggest a possible solution to improving the use of student surveys—finding ways to increase a teacher's acceptance of a student's ability to give input should increase a teacher's value of that input.

Looking over the expected changes in instruction due to student input most often cited by teachers may at first evoke a, "Well, duh," response. The list mirrors Basic Good Teaching

Practices 101. What is interesting to note, however, is that the students seem to have an understanding of and even an expectation for good practice. The teachers revealed that student opinions often pointed out teacher behaviors of which the teacher had been previously unaware or did not realize the student perception of the behavior. One teacher made the comment in a focus group discussion that he was surprised to receive several student comments about his favoritism. He said, "I had no idea they perceived some of my practices in that way. I need to work on that." For teachers, these "no-brainers" from their students served as a catalyst to jog the teachers' brains as reminders, wake-up calls, and even comforting reassurances from the fact that students were not asking for rocket-science (unless the course was rocket science), but rather just plain good teaching. Thus, the findings back up the MET Project (2010) conclusion that, "The average student knows effective teaching when he or she experiences it" (p. 5).

Teachers cited a wide variety of responses as to the *take aways* from their student surveys on Questions 2, 3, and 10. Some were very specific as the need for a calendar comment or specific project reference from Table 4. These student tidbits may hold unique power in that specific information is more readily and easily used. One teacher remarked, "The student comment section was by far the most valuable information." The more specific student comments the more teachers indicated that they knew how or what could be either changed or discussed with students. In the science department focus group discussion, one teacher specifically mentioned that reflection and changes in instruction were facilitated better by precise student feedback. Similar comments were a common thread in other focus groups. In focus group discussion notes, teachers noted also feeling the value of helpful feedback or feel validated by positive student comments.

Teacher frustrations over student evaluations were definitely a mutual source of anxiety prevalent in all the departmental discussions. Particularly disturbing to teachers was the idea that these concerns were determined to be part of the student evaluations, come as a given with the student evaluation territory. Again, apprehension over the use of student input seemed to be in continual collision with the overall positive teacher response regarding the value of student input.

With new methods for measuring the quality of teacher instruction sweeping across the nation, the possibilities for student evaluations to impact instruction were evident in this study. The poster child statement for the whole process came from this teacher comment in a focus group, "I don't think I have ever done anything that has caused more reflection including looking at test scores." Given the new educational focus on analyzing, reflecting on, and interpreting data to improve instruction, this comment shows that student feedback is data. Further research on whether or not data can be collected that demonstrates an improvement in student learning after changes made in teacher practice due to the inclusion of student voice will be powerful and perhaps the ultimate deciding factor.

Phase II. The Question 1 (*Teacher Value*) on the teacher questionnaire represented the main research question and focus. Would teachers find value in student evaluations? In Phase I, the response to teacher value had been positive and in Phase II that positive response was even higher with 89% of teachers expressing some degree of agreement level with finding value in student evaluations and 68% of these teachers indicating agreement (number 6 on Likert scale) or strong agreement (number 7 on Likert scale) and not just somewhat agree (number 5 on Likert scale). This faculty clearly indicated that they valued student input received through the student evaluations.

Interesting to note was the mean score (4.7) for Phase II came from Question 5 (*Decrease in Anxiety Level*). The question asked the teachers if their anxiety level had decreased for Phase II because of going through the process once before in Phase I. The highest concentration of responses was at "somewhat agree." But the 4.7 response does indicate a positive response in reducing anxiety by exposure to the process. Teacher concerns, however, were also noted in the focus groups as being almost identical between Phase I and Phase II. These continuing concerns indicate that the initial uncertainty over the use of student evaluations was not a matter of just having a practice round. The process of student evaluation may have become easier, reducing some anxiety, but concerns remained for teachers.

Question 3 actually had the lowest mean with a 4.6 and the highest level of variance. This question asked teachers if the student evaluations had led to a class discussion with students where teachers asked for clarification or further input. Even though the literature suggested this has a positive benefit of student surveys, not all teachers indicated they had held such discussions. If there had been an expectation for a discussion or even a clearer directive or even training on the literature, these clarifications may have raised the teachers' responses. Teachers did indicate in the focus groups that the time to have such potential discussions has been limited because of the late timing of the evaluations. There were some exceptions. In his open-ended response, one teacher stated that he had taken initiative and created a class student survey on his own and given it out between semesters to allow for changes to be made during the year based on student input. In the departmental focus groups (see Table 10), the majority of the teachers within that group shared that they did have discussions with their classes the previous year. One teacher thought that this actually led to a meaningful activity for the last few days of school.

Question 2, *changes based on instruction*, was the second highest score and came in close to the *level of value* question at a 5.8 overall mean. Overall, teachers indicated that they had made changes in their instruction based on student feedback. In fact, not even one teacher disagreed with having made changes. Some were undecided, which may mean they believe they subconsciously may have made changes, they had not purposefully made changes or given that it was the end of the year for teachers, they simply may have forgotten changes made in instruction throughout the year. Overall, however, the data suggest that student feedback had an impact on changes in actual teacher instruction. Despite their anxiety, teachers' overall responses indicated that they had valued the student feedback enough to plan and follow through with changes in instruction. The distinct correlation of the planned changes from Phase I and the actual changes made identified in Phase II provided specific evidence for the impact that student evaluations had on teacher instruction. Teachers had responded to the voice of their students.

Teachers also indicated that student comments had been more positive (54%) in Phase II as compared to Phase I. The high level of teachers having made changes to instruction may have encouraged a more positive response from students regarding teacher instruction as evidenced by the teacher comment, "This year it was nice to know that students noticed the areas that I had made an effort to change."

Question 4, *level of reflection*, decreased from Phase I to Phase II going from a mean of 6.0 (97%) to 5.1 (89%). Perhaps teacher reflection was just naturally not as high as the initial process or first time. Still 89% represents a large percentage of teachers who found the student evaluation process a stimulus for teacher reflection. Repetition of the student evaluation process did not increase but neither did it eliminate reflection. Teachers found that student input caused them to reflect on their teaching.

With all of the effort of putting collaborative and co-operative efforts into the implementation process for student evaluations, the response to Question 6, while overall positive with a 5. 1 mean still had 32% of teachers "undecided" and only 5% indicating a strong level of agreement. Directly addressing this issue in the focus groups clarified that the sticking point was the distrust of the political legislative uncertainty and the lack of the ability of teachers to collaborate in that process more than the lack of collaboration at the school level.

An unexpected additional impact on instruction was identified in the departmental focus group discussion on school wide implications of student evaluations. With 30% of teachers in Phase I and 36% in Phase II indicating that a planned change in instruction due to student evaluations was to improve communication by updating student progress on a more regular basis, this would suggest that students had commented on this in their evaluations. This information served as a catalyst for collaboration among departments on possible department solutions and even school-wide solutions.

The most common responses for suggestions to the process were again, as in Phase I, to have the student input be collected earlier or not at the very end of the year when students have just been through the grueling end of level testing and were perhaps burnt out and ready for a break. Although the intention of the administration was to conduct the student surveys earlier, the fact that testing tied up both the computers and time of administrators left them to once again put off the surveys until the very end of the year. Perhaps, polling students in February or March before end of year testing begins would be a better solution than trying for a week or two earlier in May. The research even suggested having two evaluations, one beginning in October and then another in March as the best time for feedback allowing for teacher to implement changes resulting from student evaluations during the teaching rather than after the fact.

Phase II was administered late in the year meaning that fewer classes were surveyed due to lack of time. In general, teachers had fewer classes surveyed. Teachers showed a strong preference for having more classes give input. Because some proposed student survey implementations procedures suggest only have students give input to two randomly selected teachers from their schedule, such quicker, easier methods may not provide teachers with the same value of student evaluations if the evaluations themselves are limited and may not be representative.

The quantitative data analysis for Phase II showed that all of the questions except for Question 3 were important to Question 1 or the *level of value* teachers placed on student evaluations. Since Question 3 referred to teachers *discussing the evaluations in class*, it is understandable that this happening or not happening may not have been important to the level of teacher value. Teachers who showed a more positive response for having reflected on student evaluations, made changes in their instruction, had decreased anxiety, and appreciated collaborating in the process were more likely to find a higher *level of value* in student evaluations.

The demographic data analysis, while limited partly because of the large number of variables and small population, did not show that *gender* or *years of experience* made a difference in the *level of value* teachers' placed on student evaluations. Teachers, both experienced and inexperienced, valued student input overall. New teachers might be expected to find value due to their limited experience. However, the somewhat stereotypical portrayal of experienced teachers as being too set in their ways to care about making changes was not demonstrated in this study. While only a small difference, the analysis of subject matter did indicate that teachers in STEM (Science Technology Engineering and Math) demonstrated a

slightly lower l*evel of value* of student evaluations. Perhaps, this finding would makes sense if the hard science subjects took a less emotional approach or had less concern over student concerns than the softer sciences. These subject areas might also be considered by some to be more difficult in nature and perhaps then these teachers expect students to be less qualified to rate the subjects.

Corresponding to the teachers' suggestions that some of the most useful information came from the actual student comments on the student surveys, even small or simple statements, the researcher also found this to be true within the teachers' comments on the open-ended teacher questionnaire questions. Mirroring the reflection that student comments were the most powerful data in the minds of teachers, the actual comments of teachers on the questionnaires also proved to be the most helpful to the researcher.

After pondering over the opposition of teachers to student evaluations—seemingly in direct conflict with the positive response to the value, reflection, and impact on instruction also indicated by teachers—one small, specific teacher comment sparked the initiation of an additional research phase. On a teacher reflection with an overall negative response, one teacher ended his reflection with this statement in the "*other comments*" category: "I actually like student feedback." This comment began a new journey of exploration that became the highlight of the research study.

Phase III. The researcher hunch created by a teacher comment instigated a Phase III and proved to indeed be more than a hunch. Teachers responded positively to the wording change of one question from Phase I asking in Phase III if teachers believed that students had the ability to "give feedback" on teacher instruction rather than the ability to "rate" teacher instruction. The results for this population showed a statistically significant increase in teacher acceptance of

students being qualified to give feedback on teacher instruction as compared to students being qualified to rate teacher instruction. Even with initial wording in Phase I, there was a statistical relationship between acceptance of students' ability to rate teacher instruction and a higher level of teacher value for student evaluations, suggesting that one way to increase teacher value of student input was to increase a teacher's acceptance of a student's abilities to offer input. The findings of Phase III may have found that method.

Returning to the literature, the term "rate" or even "evaluation" or "SET" (Student Evaluation of Teacher Performance) is rarely used in current literature - but those terms seem to be still prevalent in practice. While universities cling to the term "student ratings," public schools are focusing on the word "feedback." The Great Schools Project (Stefanakos, Kesselheim, & Kostin, 2013) states that student *feedback* initiatives are sweeping across the nation. Los Angeles Unified School District calls for the use of stakeholder *feedback* (2013). Boston Public Schools implemented a teacher evaluation framework that includes student *feedback (2012)*. The MET Project (2010), catalyst for the current multiple perspective approach, also uses the word *feedback* in describing their goal to include "confidential surveys to collect student *feedback* on specific aspects of a teacher's practice" (p. 8). Despite the use of the word "feedback," the implication or assumption seems to still be "evaluative," perhaps because student input is becoming part of "teacher evaluations" and, in some states, is factored or being considered as being a component into a final summative score.

The definition of "evaluate" on Dictionary.com is "to judge or determine the significance, worth, or quality of," and the definition of "rate" is "to estimate the value or worth of; appraise;" while the definition of "feedback" is "information about reactions to a performance, used as a

basis for improvement." Clearly, the task of a summative evaluation or rating is much different than that of offering formative feedback.

The study by Little, Goe, and Bell (2009) concluded that student input is calling for the student *perspective,* and Vanderbilt University (2012) suggests teachers look at student feedback as the student's perspective vs right or wrong or reliability or validity, and simply look at feedback as the student's *reactions* to instruction. The comment by one teacher during this study agreed with Vanderbilt University. Understanding a student's perspective is different from receiving a student's judgment. If student surveys were not considered or called *evaluation*s or given a numerical value within state or local reporting of teacher summative evaluation formulas, could the possibilities for teacher value be increased and thus the possibilities for a student/teacher partnership be increased? Thus, accessing, attending to, and giving students voice, which the literature suggests may be the very answer to improving student learning.

As an added by-product of giving students voice, students may be able to cultivate a 21st century marketable, real-world problem-solving skill by learning to effectively collaborate through learning how to give feedback and developing enhanced communication competence. Johnson and Wales University (2014) created a list of employability traits, titled the *Top Ten Valued Workplace Skills*. The list includes the ability to work cooperatively with a team and give helpful feedback on others' work. Instead of just handing out a survey asking for what they think, students could be trained in the art of giving specific feedback. With this kind of input, as teachers indicated that the usefulness of student comments depended on their specificity, communication and most of all learning could increase.

Training students to give feedback and helping teachers to receive feedback not only embraces the two elements necessary for communication, a giver and a receiver, but also may

hold a crucial key to teachers perceiving more value for and more effectively implementing student "feedback" surveys on teacher performance. With the initiation of an additional phase to hopefully clarify one teacher's problem with student evaluations, the researcher discovered a valuable new focus for solutions regarding student evaluations, wait, scratch that, student *feedback* on teacher instruction.

Conclusions and Recommendations

As new methods of teacher evaluations calling for student input spread across the nation, states must not respond with quick-fix measures to address teacher quality by simply imposing rigorous evaluation procedures, which may "result in further alienation and disaffection of teaching professionals" (Oon-Seng, 2012, p. 76). Strunk (2012) also warns against the rush to implement new multiple-perspective teacher evaluation policies which could undermine the potential benefits of revised teacher evaluations, advocating instead for allowing time to learn from initial implementations of teacher evaluation in a no- or low-stakes setting. This research studied and analyzed the student survey implementation process within just such a setting resulting in the following recommendations.

By exploring recommendations can be made for possible ways to improve the student survey process by reducing teacher anxiety and increasing teacher value for student surveys given in a non-threatening, collaborative partnership setting while also giving students voice in their own education by inviting feedback from the student perspective on teacher instruction.

Combining the current research found in the literature along with the conclusions of the research study, the following recommendations were complied, and they may provide value for this particular school population to improve the implementation of student surveys. The list of

recommendations may also be applicable or serve as a resource to other schools for adding student voice to teacher evaluations.

Increasing teacher value in student voice on teacher performance.

- Refer to student input as feedback rather than ratings, rankings, or evaluations.
- Do not use the term SET (Student Evaluation of Teacher). Use term SFT (Student Feedback for Teacher).
- Use student feedback surveys as a method for accessing formative student feedback, allowing teachers an opportunity to adjust or inform instruction.
- Train teachers, students, parents, and other stakeholders in the role and value of giving students voice.
- Encourage teachers to train students in the principles of feedback throughout the year by incorporating self- reflections and peer reviews into classwork assessments.
- Have teachers discuss ways to further embed regular methods for obtaining student feedback into instructional techniques, after units, projects, tests, etc.
- Discuss other ways teachers may be able to give students meaningful roles in classroom collaborative efforts and on-going feedback to improve student learning and teacher instruction.
- Have teachers include references during instruction to changes the teacher has made due to prior student feedback throughout the year.
- Hold trainings for the faculty on how to interpret and use student feedback before teachers have access to the results. Focus teachers on feedback as accessing the

students' perspective rather than determining the correctness or the right or wrong nature of the feedback.

- Have teachers ask questions and clarify input with students in a safe classroom discussion of student survey feedback.

Improving the implementation process for student voice on teacher performance.

- Form a collaborative school implementation team. The team could be a representative from each department or grade level and could include students.
- Create projected timelines for implementation. Two student surveys, one in October and one in March or could be administered to make surveys timely. If time limitations are present, teachers can administer their own October surveys and an administrator can administer March surveys. October student feedback give teachers the opportunity to discuss evaluation results with students for clarification and give students the opportunity to view their role as formative partners in the midst of the process rather than just an afterthought at the end.
- Develop a student survey with district and state requirements that allows for an inclusion of individual school needs and not just a review of research-based survey questions. The implementation process including the survey instrument and type can be discussed collaboratively.
- Allow for teacher input to decide on constructs or values of the classroom and teacher instruction to be measured that would be the most beneficial to teachers and the most applicable for student input.
- Use student-friendly language and positive phrasing in questions, measuring one construct at a time

- Consider adding a student self-reflection piece to add to the student survey to increase student awareness of their role in the partnership.
- Collect feedback from students, teachers, administrators, parents, and district leaders on the initial draft and the implementation ideas. School community councils or PTA's can be a source of parent feedback. Student councils as well as polling students from a variety of different school groups or classes can be a method for receiving student input. Clarifying student interpretation of questions to original intent can be an important part of these discussions. Collaboration or departmental meetings can help to facilitate faculty collaborative input.
- Plan for training the student body on the purpose and role of feedback and the value of student/teacher partnership before administering the survey. Define constructive feedback. Constructive (adjective) is serving a useful purpose; tending to build up. Synonyms: useful, helpful, and productive. Feedback (noun) information about a person's performance of a task, used as a basis for improvement.
- Model, demonstrate, practice, address what if's, discuss sample questions, show examples, teach and illustrate the value of being specific for students.
- Explain that an evaluation containing crude comments or personal attacks is more of a reflection on the person writing the evaluation than the one being evaluated and will be removed from surveys shown to teachers.
- Use counselors and other office staff to help in training. The training presentation could be a live video feed or a filmed presentation shown in classrooms. Grade

level meetings or assemblies may also be used. Once this becomes part of regular practice, only incoming students will need to be trained.

- Make a connection to Common Core thinking and writing and feedback as a marketable skill. Frame the comment section in student surveys. Explain to students that these frames are prompts and should be addressed as they would a writing prompt: staying on prompt, making statements, and backing the statement up with specific details. For example, the student open-ended comments section rather than just saying, "Comments," could use one of the following prompts or a combination.
    - What was the best thing about this class?
    - What are some specific suggestions you have to help your teacher improve?
    - What is the most important thing you learned in this class?
    - What could your teacher do to help increase learning?
    - What parts of this class helped you learn best?
    - What specific advice would you give to help your teacher improve your learning in this course?
- Give administrators accountability by including a summative goal setting discussion with teachers, including at least one based on formative student feedback. Progress toward these formative goals could be used on summative teacher evaluations. Thus, accessing the student perspective for the 360 model without making the process only punitive for teachers.

- Discuss the student feedback within collaboration teams or focus groups. Teams can also discuss the student feedback with administrators so that school wide or departmental patterns may also inform additional school-wide or department or grade level goals.
- Consider adding additional student surveys for administrators, coaches, club advisors, or other specific school needs.
- Evaluate the process at the end of the first year again involving all stakeholders to make any needed changes. The process then can be reviewed as needed as it becomes embedded in teacher formative practice and the school improvement plan.

As these recommendations and additional student survey methods are implemented in other contexts, new insights may be studied and shared. Future research could address specific methods for training students in the art of giving feedback and for training teachers in interpreting and incorporating student feedback. The determination of 360-model teacher evaluation formulas and how to factor in student feedback is still a piece of the student feedback puzzle and an area for further study. The actual impact of including student voice on teacher performance as to changes in measured levels of learning and impact on school culture and climate will be the work of future research. While more research needs to be done, this study questions the rush to punitive evaluation practices, but also highlights teachers' willingness to partner with their students, find value in their voice, and adjust instruction to meet their needs. This study demonstrates that student feedback of teacher instruction has the potential, if conducted in a positive not punitive approach, to access student voice and create change: a win-win student/teacher partnership that improves teacher instruction and student learning.

References

Abrami, P., Leventhal, L., & Perry, R. (1982). Educational seduction. *Review of Educational Research*, *52*(3), 446-464.

Adams, J. (1997). Student evaluations: The ratings game. *Inquiry*, *1*(2), 10-16.

Alaska Statute Title 14, Chapter 20, Section 149. (2010). *Employee evaluation.* Retrieved from http://www.legis.state.ak.us/basis/sta tutes.asp?title=14

Aleamoni, L. (1999). Student ratings myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*(2), 153-166.

Andrew, S., Salamonson, Y., & Halcomb, E. (2008). Integrating mixed methods data analysis using NVivo: an example examining attrition and persistence of nursing students. *International Journal of Multiple Research Approaches*, *2*(1), 36-43.

Barile, J., Donohue, D., Anthony, E., Baker, A., Weaver, S., & Henrich, C. (2012). Teacher–student relationship climate and school outcomes: Implications for educational policy initiatives. *Journal of Youth and Adolescence*, *41*(3), 256-267.

Beaudoin, N. (2005). *Elevating student voice*. Larchmont, NY: Eye on Education.

Bergmark, U., & Kostenius, C (2009). Listen to me when I have something to say: students' participation in research for sustainable school improvement. *Improving Schools, 12*(3), 249-260.

Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Research. 31*(8), 18-20.

Bill & Melinda Gates Foundation. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching Project.* Retrieved from http://www.metproject.org/ downloads/Preliminary_Findings-Research_Paper.pdf

Blanchard, B. E. (1968). Illinois ratings of teacher effectiveness. *The New Campus, Teachers College Association for Extension on Field Services*, *21*, 28-30.

Boser, U., & Rosenthal, L. (2012, July 10). To improve schools, we should listen to students. US News and World Report. Retrieved from http://www.usnews.com/opinion/articles/2012/07/10/to-improve-schools-we-should-listen-to-students

Boston Student Advisory Council. (Spring 2012). We are the ones in the classroom-ask us: Student voice in teacher evaluation. *Harvard Educational Review, 82*(1)*,* 153-162,166.

Brenchley, C. (2011, March 25). *Top 5 things we heard at the National Youth Summit*. [Web log message]. Retrieved from http://www.ed.gov.erl.lib.byu.edu/blog/2011/03/top-5-thingswe-heard-at-the-national-youth-summit/

Bryan, R. C. (1966). The teacher's image is stubbornly stable. *The Clearing House, 40*, 459-461.

Calkins, S., & Micari, M. (Fall 2010). Less-than-perfect judges: Evaluating student evaluations. *NEA Journal of Higher Education.* Retrieved from http://www.nea.org/assets/img/PubThoughtAndAction/TA10CalkinsMicariR.pdf

Cambridge Education. (2013). Retrieved from http://tripodproject.org/services-offerings/survey-assessments

Camp, N. J., & Slattery, M. L. (2002). Classification tree analysis: a statistical tool to investigate risk factor interactions with an example for colon cancer. *Cancer Causes and Control, 13* (9), 813-823.

Cavanaugh, S. (2014). Companies honing tools to survey students about teachers. *Education Week.* Retrieved from http://www.edweek.org/ew/articles/2014/08/27/02surveys.h34.html

Center for American Progress and the Institute for America's Future. (2005). *Getting Smarter, Getting Fairer.* Washington, DC: Center for American Progress and the Institute for America's Future.

Charmaz, K. (1990). Discovering chronic illness: Using grounded theory. *Social Science Medicine*, *30*, 1161-1172.

Cho, J., & Trent, A. (2006). Validity in qualitative research revisited. *Qualitative Research, 6*(3), 319-340.

Chulkov, D., & Van Alstine, J. (2012). Challenges in designing student teaching evaluations in a business program. *International Journal of Educational Management*, *26*(2), 162 –174.

Cohen, P. (1990). Bring research into practice. In M. Theali & J. Franklin (Eds). *Student Ratings of Instruction: Issues for Improving Practice: New Directions for Teaching and Learning.* San Francisco, CA: Jossey-Bass.

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1), 3.

Committee for Economic Development. (1985). *Investing in our children: Business and the Public Schools.* New York, NY: Carnegie Group of New York.

Cook-Sather, A. (2002). Authorizing students' perspectives: Toward trust, dialogue, and change in education. *Educational Researcher*, *31*(4), 3-14.

Creswell, J. W. (2002). *Educational Research*. Saddle River, New Jersey: Pearson Education.

Daley, J. (1999). Trust, but verify: Toward fairer student evaluations. *NEA Journal of Higher Education.* Retrieved from http://www.nea.org/assets/img/PubThoughtAndAction/TAA_99Fal_09.pdf

Davidoff, S. H. (1970). The development of an instrument designed to secure student assessment of teaching behaviors that correlate with object measures of student achievement. Philadelphia School District. ED 039 170.

DeMarrais, K., & Suggs, C. (2011). *Critical contributions: Philanthropic investment in teachers and teaching*. Atlanta: University of Georgia and Kronley and Associates. Retrieved at http://kronley.com/documents/CriticalContributions.pdf

Dillon, S. (2010, December 7). Top scores from Shanghai stun educators. *New York Times.* Retrieved from http://www.nytimes.com/2010/12/07/education/07education.html?_r=1&pagewanted=1

Doyle, M., & Feldman, J. (2006). Student voice and school choice in the Boston Pilot High Schools. *Educational Policy*, *20*(2), 367-398.

Dufour, R., & Eaker, R. (1998). *Professional Learning Communities at Work*. Bloomington, ID: National Education Service.

Duncan, A. (2011, February). *Keynote speech*. Speech presented at Voices in Action National Youth Summit, Washington, DC.

Ebmeier, H., Jenkins, R., & Crawford, G. (1991). The predictive validity of student evaluations in the identification of meritorious teachers. *Journal of Personnel Evaluation in Education*, *4(*5), 341-357.

Erickson, F. (1987). Transformation and school success: The politics and culture of school ethics. *Educational Administration, 2*(1): 1–8.

Erlandsen, D., Harris, E., Skipper, B., & Allen, S. (1993). *Doing Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Sunshine Review. (2012). "Faculty evaluations and public records." Retrieved from http://sunshinereview.org/index.php/Faculty_evaluations_and_public_records#cite_note-11

Farreras, I. (2012). The effect of faculty self-promotion student evaluations of teaching. *College Student Journal, 46*(2), 314-322.

Feuer, A. (2009). Student board members make a difference. *Education Digest*, *74*(9), 17-19.

Fielding, M. (2001). Students as radical agents of change. *Journal of Educational Change*, *2*(2), 123-141.

Fielding, M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal, 30*(2), 295-311.

Fish, S. (2010, June 21). Deep in the heart of Texas. *New York Times Opinion Page.* Retrieved from http://opinionator.blogs.nytimes.com/2010/06/21/deep-in-the-heart-of-texas

Fish, S. (2010, June 21). Student Evaluations Part Two. *New York Times Opinion Page.* Retrieved from http://opinionator.blogs.nytimes.com/2010/06/28/studentevaluationsparttwo

Flinn, V. (1932). Teacher rating by pupils. *Educational Method*, *11*, 290-294.

Fullan, M. (2001). *The new meaning of educational change*. (3rd Ed.). New York: Teachers College Press.

Glazier, J. (2005). Teaching through a positional lens: Recognizing what and who we privilege in our practice. *Teaching Education, 16*(3), 241-243.

Goodlad, J., Soder, R., & Sirotnik, K. (Eds.). (1990). *The moral dimensions of teaching*. San Francisco, CA: Jossey-Bass Publishers.

Hart, R. (1992). Children's participation: From tokenism to citizenship. UNICEF Innocenti Essay 4. Retreived from http://www.unicef-irc.org/publications/pdf/childrens_participation.pdf

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research*, *3*(3), 118-126.

Holcomb, E. (2004). Getting excited about data: Combining people, passion, and proof in education. *Educational Researcher*, *31*(4), 3-14.

James, R., & Rottman, C. (2009). Struggling for democracy: Administrative communication in a diverse school context. *Educational Management, Administration, and Leadership*, *37*(4), 473-496.

Johnson and Wales University. (2014). *Top ten valued workplace skills*. Retrieved at http://www.jwu.edu/uploadedFiles/Documents/Alumni/JWUTopWorkSkills.pdf

Kane, T. (2013). Ask the students. *Brookings Institute*. Retrieved from www.brookingsinstitute.edu/blogs/brown-center-chalkboard/posts/2013/04/10-teachers-evaluations-kane

Kauchak, D., Peterson, K., & Driscoll, A. (1985). An interview study of teachers' attitudes toward teacher evaluation practices. *Journal of Research and Development in Education, 19*, 32-37.

Kington, A., Sammons, P., Day, C., & Regan, E., (2011). Stories and Statistics: describing a mixed method study of effective classroom practice. *Journal of Mixed Methods Research, 5*(2), 103-125.

Kitzinger, J. (1995). Qualitative research: introducing focus groups. *British Medical Journal, 3*(11), 299-301.

Knight, F. B. (1922). *Qualities rated to success in teaching*. New York City: Teachers College, Columbia University.

Kozub, R. (2010). Relationship of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness in business schools. *American Journal of Business Education*, *3*, 33-41.

Kratz, H. E. (1896). Characteristics of the best teachers as recognized by children. *Pedagogical Seminary, 3*, 413- 418.

Kulik, J. (2001). Student ratings: validity, utility, and controversy. *New Directions in Institutional Research*, *109*, 9-25.

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

Lett, J. (1990). Emics and etics: Notes on the epistemology of anthropology. In T.N. Headland, K.L. Pike, & M. Harris (Eds.). *Emics and etics: The insider/outsider debate. Frontiers of anthropology,* Newbury Park, CA: Sage Publications.

Levin, B. (2000). Putting students at the centre in education reform. *International Journal of Educational Change, 1*(2), 155-172.

Lewin, K. (1951). *Field theory in social science; Selected theoretical papers.* D. Cartwright (Ed.). New York: Harper & Row.

Lewis, K. G. (2001). Making sense of student written comments. *New Directions for Teaching & Learning*, *2001* (87), 25.

Light, U. L. (1930). High school students rate teachers. *The School Review*, *38*, 28-32.

Lincoln, Y., & Guba, E. (1985). *Naturalistic Inquiry*. Beverly Hill, CA: Sage.

Little, O., Goe, L., & Bell, C. (2009). A practical guide for enhancing teacher effectiveness. *National Comprehensive Center for Teacher Quality*.

Los Angeles Unified School District. (2012). Educator Performance Plan. Retrieved from http://talentmanagement.lausd.net

Lunenberg, F., & Ornstein, A. (2004). *Educational administration*. Belmont, CA: Wadsworth/Thompson Learning.

Making Sense of Student Evaluation Feedback. (2012). Vanderbilt University: Center for Teaching and Learning. Retrieved from http://cft.vanderbilt.edu/teaching-guides/reflecting/student-evaluations

Manatt, R. (1997). The 360-degree feedback process for professional evaluation. The School Administrator. Retrieved from http://www.aasa.org/SchoolAdministratorArticle.aspx?id=15598

Marsh, H. & Hocevar, D. (1991). The multidimensionality perspective on students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching and Teacher Education,* 7, 9-18.

Maslow, A. H. (1943). A theory of human motivation, *Psychological Review, 50*(4), 370-396.

Mason, J. (2002). *Qualitative researching*. London, England: Sage Publications.

Masters, J. R., & Weaver, W. G. (1977). *The development of a student observation of teachers instrument for use in high schools.* Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

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

McPherson, M., Jewell, R., & Kim, M. (2009). What determines student evaluation scores? *Eastern Economic Journal, 35,* 37-51.

Medley, D. M., & Mitzel, H. E. (1959). Some behavioral correlates of teacher effectiveness. *The Journal of Educational Psychology*, *50*, 239-246.

Miles, B., & Huberman, A. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage Publications.

Mitra, D., & Gross S. (2009). Increasing student voice in high school reform. *Educational Management Administration and Leadership*, *37*, 522.

Mitra, D. (2003). Student voice in school reform: Reframing student-teacher relationships. *McGill Journal of Education*, *38*(2), 289-304.

Mitra, D. (2005). Adults advising youth: Leading while getting out of the way. *Educational Administration Quarterly*, *41*, 520-553.

Mitra, D. (2006). Increasing student voice and moving toward youth leadership. *Prevention Researcher*, *13*(1), 7-10.

Moldof, E. P. (1993). Schools and strategic planning: Using a traditional business tool. *School Business Affairs, 59*(2), 21-24.

Morgan, J. (2014). Classification and regression tree analysis. *Technical Report Boston University School of Public Health*. Retrieved from http://www.bu.edu/sph/files/2014/05/MorganCART.pdf

Moore, S., & Kuol, N. (2005). Punitive bureaucratic tool or valuable resource? Using student evaluations to enhance your teaching. In R. O'Neill, S. Moore, & B. McMullin (Eds.), *Emerging Issues in the Practice of University Learning and Teaching*. Dublin, Ireland: All Ireland Society for Higher Education (AISHE).

Naftulin, D. H., Ware, J. E., Jr., & Donnelly, E. A. (1983). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, *75*, 138-149.

National Center for Educational Statistics. Retrieved from http://nces.ed.gov/ccd/schoolsearch/school_detail.asp

Obama, B. (2009). Obama outlines five educational top priorities. Retrieved from www.pbs.org/newshour/updates/education/.../education_03-10.html

O'Malley, M. D., Renshaw, T. R., Ritchey, K. M., & Furlong, M. J. (2011). Gauging the system: Trends in school climate measurement and intervention. In S. Jimerson, M. Mayer, A. Nickerson, & M. Furlong (Eds.), *Handbook of school violence and school safety: From research to practice* (2nd ed.). New York: Routledge.

Onwuegbuzie, A., Daniel, L., & Collins, K. (2009), A meta-validation model for assessing the score-validity of student teaching evaluations, *Quality and Quantity*, *43,*197-209.

Oon-Seng, T. (2012). Singapore's holistic approach to teacher development. *Phi Delta Kappan*, *94*(3), 76-77.

Ory, J. C. (2001). Faculty thoughts and concerns about student ratings. *New Directions for Teaching & Learning*, *2001*(87), 3-15.

Owens, R. G. (1998). *Organizational behavior in education*. Needham Heights, MA: Allyn and Bacon.

Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd Ed.). Thousand Oaks, CA: Sage Publications.

Pearson, M. (2012). Wins, losses, and draws in Chicago school strike. *CNN*. Retrieved from http://www.cnn.com/2012/09/19/us/illinois-chicago-teachers-strike/index.html

Peck, R. F., Olsson, N. G., & Green, J. L. (1978, August). *The consistency of individual teaching behavior*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada.

Perry, P., Hladkyj, S., Pekrun, R. H., Clifton, R. A., & Chipperfield, J. G. (2005). Perceived academic control and failure in college students: a three-year study of scholastic attainment. *Research in Higher Education*, *46*(5), 535-569.

PISA (Program for International Student Assessment). (2013). What students know and can do. Retrieved from http://www.oecd.org/dataoecd/54/12/46643496.pdf

Policy Statement on Teacher Evaluation and Accountability. (2011). National Education Association Annual Meeting and Representative Assembly. Retrieved from http://www.nea.org/grants/46084.htm

Public Education Employment Reform. (2012). UT-H.B.64, State of Utah General Legislative Session 2012.

Remmers, H. H. (1934). Reliability and halo-effect of high school and college students' judgments of their teachers. *Journal of Applied Psychology, 18*, 619-630.

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

Ripley, A. (2013). *The Smartest Kids in the World and How They Got That Way*. New York, NY: Simon and Schuster.

Ritter, G. (2010). The delicate task of developing an attractive merit pay plan for teachers. *Phi Delta Kappan*, *91*(8), 32-37.

Rothstein, J., & Mathis, W. (2013). Review of two culminating reports from the MET project. *National Education Policy Center*. Retrieved from http://nepc.colorado.edu/thinktank/review-MET-final-2013

Rudduck, J., & Demetriou, H. (2003). Student perspectives and teacher practices: The transformative potential. *McGill Journal of Education*, *38*(2), 274-287.

Sanford, H. (2013, May). Students speak: teachers in Lee County, FL embrace student feedback to improve instruction. *The NEA Foundation Issue Brief*. Retrieved from http://www.neafoundation.org/public/uploads/1368553590NEAF_IssueBrief8_May2013.pdf

San Francisco Youth Commission. (2009). Annual report. Retrieved from http: //www.sfbos.org

Santorico, S., & Shan, M. (2010). Tutorial for the R statistical package. Retrieved from http://math.ucdenver.edu/RTutorial/Tutorial_guide.pdf

Schwab, R. L. & Iwanicki, E. F. (1988). *Merit pay: Stressor or motivator*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana.

Seldin, P. (1999). *Changing Practices in Evaluating Teaching: A practical guide to improved faculty performance and promotion/tenure decisions*. Boston, MA: Anker.

Shah, N. (2013). States use school score cards to target climate problems. *Education Week*, *32*(26), 1-12.

Sinagub, J., Vaughn, S., & Schumn, S. (1996). *Focus Group Interviews in Education and Psychology*. Thousand Oaks, CA: Sage Publications, Inc.

Smith, T. (September 16, 2012). Teacher evaluation dispute echoes beyond Chicago.

NPR. Retrieved from http://www.npr.org/2012/09/13/160247078/teacher-evaluation-dispute-echoes-beyond-Chicago

Smyth, J., & McInerney, P. (2006). *Teachers in the middle: Reclaiming the wasteland of young adolescence.* New York: Peter Lang Publishing.

Smyth, J. (2006). When students have power: Student engagement, student voice, and possibilities for school reform. *International Journal of Leadership in Education*, *9*(4), 285-298.

Spencer, K., & Schmelkin, L. (2002). Student perspectives of teaching and its evaluation. *Assessment and Evaluation in Higher Education*, 27(5), 397-409.

Spindler, G., & Spindler, L. (1987). *Interpretive ethnography of education: At home and abroad*. Hillsdale, NJ: Lawrence Erlbaurn Associates.

Spradley, J. (1980). *Participant observation*. New York: Holt, Rinehart and Winston.

Strauss, A., & Corbin, J. (1998). *Review of Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory*. London: Sage Publications.

Stefanakos, M., Kesselheim, C., & Kostin, M. (2013). Using student feedback to improve instruction. *Great Schools Project.* Retrieved from http://www.greatschoolspartnership.org/using-student-feedback-improve-instruction

Strunk, K. (2012). Lessons learned: Los Angeles Unified School District. *Phi Delta Kappan*, *94*(3), 47-51.

Sullivan, J. (Spring 2012). A collaborative effort: Peer review and the history of teacher evaluations in Montgomery County, Maryland. *Harvard Educational Review 82*(1), 142-15.

Svinicki, M. D. (2001). Encouraging your students to give feedback. *New Directions for Teaching and Learning, 2001*(87), 17-24.

Tips and tools: Coding qualitative data. (2012). *Center for Evaluation and Research*. Retrieved from http://tobaccoeval.ucdavis.edu/documents/Tips_Tools_18_2012.pdf

Toch, T. (2012). The fallacy of local control. *Phi Delta Kappen. 93*(5), 66-67.

Trout, P. A. (1997). What the numbers mean: Providing a context for numerical student evaluations of courses. *Change*, 29, 25-30.

Unger, C. (2003). Why listen to students? *New Horizons of Learning* (John Hopkins School of Education). Retrieved from http://education.jhu.edu/PD/newhorizons/student-voices-12/listen

University of Wisconsin Handbook. (2012). Retrieved from http://www.ls.wisc.edu/handbook/ChapterFive/chV-14.htm

U. S. Department of Education (2012). RESPECT Project, Retrieved from http://www.ed.gov/news/press-releases/obama-administration-seeks-elevate-teaching-profession-duncan-launch-respect-pro

USOE. (2014). *Common Core resources*. Retrieved from www.utah.schools.gov/core

US News/Education. (2012). Retrieved from http://www.usnews.com/education/best-high-schools/utah/districts

Walker, T. (2011). *RA delegates adopt NEA policy statement on teacher evaluation and accountability.* NEA. Retrieved from http://www.nea.org/grants/46084.htm

Weick, K. E. (1987). Substitutes for strategy. In D. J. Teece (Ed.), *The competitive challenge: Strategies for industrial innovation and renewal* (pp. 221-234). Cambridge, MA: Ballinger Publishing.

Williams, L., Cate, L., & O'Hair, M. (2009). "The boundary spanning role of democratic learning communities." *Educational Management Administration and Leadership*, *37*, 452-472.

Woessmann, L. (2011). Merit pay international. *Education Next, 11*(2), 72-77.

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

Yonezawa, S., & Jones, M. (2009). Student voices: Generating reform from the inside out. *Theory Into Practice, 48*, 205-212.

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

Zinth, J. D. (2010). Teaching quality: Evaluation and effectiveness. Denver, CO*: Education Commission of the States.* Retrieved from www.ecs.org/clearinghouse/86/21/8621.pdf

Appendix A

Consent to Be a Research Subject

Introduction

This research study is being conducted by **Brenda Burr,** a doctoral student at Brigham Young University/ Educational Leadership Department to understand the process of student evaluation of teachers and how teacher anxiety might be decreased through a collaborative process, teachers' perspectives and value of these evaluations, and how student evaluations may inform teacher instructional practice. You have been selected as a faculty member of Salem Hills High School to participate in an archival study of the administration of Phase I and Phase II of school wide student evaluations consisting of the results of teacher questionnaires and focus group discussions.

Procedures

If you agree to participate in this research study, the following will occur:

- You will also be asked to give permission for the release of your teacher questionnaire response and notes from any focus groups you may have participated in for both Phase I and Phase II of the school archival study of the implementation of student evaluations.
- You will be asked to view and offer feedback on the completed research findings.

Risks/Discomforts

There are minimal risks for participation in this study.

Benefits /Compensation

There will be no direct benefits or compensation. It is hoped, however, that through your participation researchers may gain new insights into student evaluations of teachers and the results may be able to inform the implementation of teacher evaluations as a tool for improving teacher instructional practice.

Confidentiality

The actual student evaluation results of you as a teacher will be stored in an on-line password protected program managed by the school principal. Only you and your administrator will have access to the evaluation results. The student evaluations of teachers will not be used in this study. The research data from individual teacher questionnaires and focus group discussions will be kept confidential and only the researcher and principal will have access to the data. Your name will not be used in the research. If quotes are used from teacher questionnaire or focus group discussion, no names will be attached. Names of teachers or students will not be used.

Participation

Participation in this research study is voluntary. You may choose to decline or excuse yourself from any question you feel uncomfortable in answering. You may withdraw from the research at any time without any work-related consequences.

Questions about Your Rights as Research Participants

If you have questions regarding your rights as a research participant, contact Brenda Burr – brenda.burr@nebo.edu or (801) 362-0719 or IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Participant Name (Printed):

Signature Date:
Researcher: Brenda Burr Signature: Date:

Appendix B

Second Round Student Evaluation Survey

1. My teacher uses class time effectively.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

Comments

2. When I am confused, my teacher helps me get straightened out.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

3. My teacher enjoys teaching.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

4. I know what my teacher wants me to learn each class period.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree

- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

5. My teacher encourages active student participation in the classroom.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

6. My teacher keeps me updated on my progress.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

7. My teacher responds to all students respectfully.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

8. My teacher demonstrates good knowledge of the subject material of this class.

- 1 Strongly Agree
- 2 Agree

- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

9. My teacher sincerely cares about me as a person.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

10. My teacher manages classroom behavior.

- 1 Strongly agree
- 2 Agree
- 3 Somewhat agree
- 4 Neither agree or disagree
- 5 Somewhat disagree
- (•) 6 Disagree
- 7 Strongly disagree

11. I spend time studying outside of class.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

12. I completed all of the homework assignments for this class.

- 1 Strongly Agree

- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

13. I attend class regularly.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

14. I put forth my best effort in this class.

- 1 Strongly Agree
- 2 Agree
- 3 Somewhat Agree
- 4 Neither Agree Nor Disagree
- 5 Somewhat Disagree
- 6 Disagree
- 7 Strongly Disagree

18. What did you like most about this class?

19. How could this class be improved.

.

Prev Done

Appendix C

Phase I Teacher Questionnaire on Student Evaluations

No Name Needed

Circle your response for first four answers

1. Overall, I found the student survey experience valuable.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

2. I will make changes in my instruction due to student input from the survey

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

3. The student survey facilitated reflection on my practice.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

4. I believe students are qualified to rate my instruction.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

Please comment on the following questions.

1. What insights did you gain from the data? Any information valuable from comparing lower scores to higher scores?

2. Did you gain any valuable information from student comments section? Explain?

3. Now that you have seen the results, would you like to add or take out any questions?

4. Will you make any changes in instruction because of the results of your student evaluations? What specific change will you make?

5. Overall what did you find the most valuable from the student evaluations?

6. What are your biggest concerns with student evaluations?

7. Do you have ideas on how to implement or administer student evaluations other than the way these were done?

8. Overall, do you see a value in student evaluations?

9. Any other information or comments

Appendix D

Phase I Collaborative Focus Group Discussion

Department _____ Number of department present _____

*Focus group leader (Department Head), please write down comments concerns, or insights from your discussion about the questions each teacher answered on the survey.

1. What insights did you gain from the data? Any information valuable from comparing lower scores to higher scores?

2. Did you gain any valuable information from student comments section? Explain?

3. Now that you have seen the results, would you like to add or take out any questions?

4. Will you make any changes in instruction because of the results of your student evaluations? What specific change will you make?

5. Overall what did you find the most valuable from the student evaluations?

6. What are your biggest concerns with student evaluations?

7. Do you have ideas on how to implement or administer student evaluations other than the way these were done?

8. Overall, do you see a value in student evaluations?

9. Any other information or comments

Appendix E

Phase II Teacher Questionnaire

No Name Needed

Years of Teaching Experience in Public School 1-3_____ 4-6 _____7 or above _____

Gender (M or F) _____

Subject Taught _____

Grade Level Taught _____

1. Overall, I found the student survey experience valuable.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

2. I made changes in my instruction due to student input from the survey.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

3. I used input from student evaluations last year to facilitate classroom discussions this year.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

4. Overall, student evaluations have caused me to reflect on my teacher practice.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

5. My level of anxiety has been reduced after participating in the student evaluation process.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

6. The collaborative nature allowed in the implementation of student evaluations increased my level of support for the process.

1 Strongly disagree 2 Disagree 3 Somewhat disagree 4 undecided 5 Somewhat agree 6 Agree 7 Strongly agree

Please comment on the following questions.

1. How did your evaluation results differ from last year?

2. Identify changes you noticed due to question revisions from initial evaluations?

3. List specific changes made in instruction this year based on last year's evaluations.

4. What additional changes will you make in instruction because of the results on this year's student evaluations?

5. What did you find the most valuable from the student evaluations (either time)?

6. What did you find the least valuable from the student evaluations (either time)?

7. What is still your biggest concern with student evaluations?

8. What did you notice by not having as many classes give feedback this year?

9. What ideas do you have on how to make additional changes to the implementation process?

10. Any other questions or comments?

Appendix F

Phase II Collaborative Focus Group

Department Heads_____

Number of faculty present _____

1. How did your evaluation results over all as a department differ from last year?

2. What information was gained from student comments section that could be helpful to the department as a whole?

3. Which questions would the department to change or revise?

4. Identify changes you noticed as a department due to question revisions from initial evaluations?

5. What changes were made in the department as a whole based on last year's evaluations?

6. How did your department inform students about changes made in instruction dues to last year's evaluation?

7. What additional changes will the department make in instruction because of the results of this year's student evaluations?

8. What did the department see as the most valuable from the student evaluations?

9. What did the department see as the least valuable from the student evaluations?

10. What are still the department's biggest concerns with student evaluations?

11. What concerns were reduced for the department by changes made in the implementation process from the initial to the second round?

12. What ideas do you have as a department on how to make additional changes to the implement process?

13. Explain the value your department sees or does not see in student evaluations?

14. Any other information or comments

Appendix G

Quantitative and Qualitative Data Spreadsheet for Phase II

quant1	quant2	quant3	quant4	quant5	quant6	quant7	quant8	subject
6	7	5	7	2	4	B	F	FA/Hum
6	7	7	7	6	6	M	F	Other
6	6	2	7	7	6	V	M	FA/Hum
6	6	5	6	6	6	V	F	FA/Hum
6	5	4	5	2	4	V	M	STEM
7	6	3	7	6	5	V	F	Other
5	5	5	5	2	2	V	M	STEM
6	6	6	6	4	5	V	M	STEM
2	5	1	5	6	5	V	M	STEM
6	6	6	6	4	6	V	M	STEM
6	5	6	7	5	5	V	M	FA/Hum
5	5	2	6	4	4	V	M	FA/Hum
7	6	5	7	4	6	M	M	FA/Hum
6	6	2	5	4	4	B	M	Other
6	5	5	3	5	5	V	M	FA/Hum
4	6	5	6	3	6	V	M	FA/Hum
7	6	5	7	6	6	V	M	FA/Hum
7	4	3	6	4	6	B	F	FA/Hum
7	7	7	7	1	4	V	M	STEM
6	7	7	6	4	5	B	F	STEM
6	5	2	7	5	6	V	F	STEM
7	6	6	7	4	7	B	M	FA/Hum
6	5	5	7	7	4	B	F	STEM
6	5	2	5	4	4	B	F	STEM
6	7	7	7	6	6	M	M	Other
4	5	5	4	4	5	V	M	FA/Hum
6	7	6	7	6	6	V	M	Other
7	6	4	7	7	7	M	F	FA/Hum
6	5	3	6	5	4	V	F	FA/Hum
6	6	3	6	6	5	V	M	FA/Hum
5	5	2	4	4	4	V	F	FA/Hum
4	4	4	4	4	4	V	M	STEM
6	6	5	7	6	4	V	F	FA/Hum
6	5	6	6	5	6	V	F	FA/Hum
6	7	7	7	5	6	B	M	Other
6	5	6	7	6	6	M	F	Other
5	5	5	5	4	4	M	M	Other

qual1	qual2	qual3	qual4	qual5	qual6	qual7	qual8	qual9
more	none	classroom	classroom	feedback	none	punitive	not repres	other
more	none	classroom	hands on	feedback	none	punitive	none	none
more	none	letting	timely upd	other	none	none	none	none
more	more	workload	timely upd	validation	none	other	none	none
same	none	student	help all	feedback	irrelevant	validity	not repres	none
more	more	variety	self	feedback	irrelevant	punitive	none	other
more	none	variety	hands on	other	student	none	none	none
more	none	none	none	feedback	none	none	not repres	none
same	none	none	timely upd	other	conflicting	none	none	other
more	more	letting	continue c	feedback	student	other	none	none
same	more	communic	hands on	feedback	conflicting	none	not repres	none
same	none	letting	timely upd	other	none	none	not repres	none
same	none	workload	continue c	feedback	student	none	not repres	none
more	more	classroom	none	validation	none	other	fewer	none
same	none	none	none	other	none	none	none	none
more	more	letting	classroom	other	none	punitive	none	earlier
same	none	none	timely upd	other	irrelevant	none	none	none
more	more	workload	timely upd	feedback	student	validity	not repres	other
same	none	workload	add	reflection	none	none	fewer	none
more	more	student	continue c	other	none	punitive	not repres	none
more	none	student	continue c	reflection	irrelevant	none	not repres	none
same	none	none	help all	reflection	student	none	not repres	none
more	none	classroom	continue c	other	none	none	none	none
same	none	letting	classroom	other	student	none	not repres	earlier
more	none	letting	none	other	none	none	none	none
more	quicker	communic	add	other	conflicting	punitive	none	none
same	none	communic	continue c	other	none	other	not repres	earlier
more	none	classroom	timely upd	reflection	none	none	not repres	none
more	more	letting	timely upd	validation	student	none	not repres	none
more	none	preparatio	timely upd	feedback	student	other	not repres	none
same	quicker	none	none	other	none	punitive	none	none
more	none	none	self	other	none	none	not repres	none
more	more	preparatio	help all	validation	student	none	not repres	earlier
more	none	student	add	validation	student	punitive	none	none
same	none	none	help all	feedback	student	none	none	none
same	none	classroom	none	validation	none	none	none	none
same	none	none	none	validation	student	none	none	none