Motivational Forces in a Growth-Centered Model of Teacher Evaluation

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Education

by

Nicholas Aron Bruski

2012
ABSTRACT OF THE DISSERTATION

Motivational Forces in a Growth-Centered Model of Teacher Evaluation

by

Nicholas Aron Bruski

Doctor of Education

University of California, Los Angeles, 2012

Professor Christina Christie, Chair

This paper presents the results of a study that explored the effects of using an action research process to examine and develop a system of teacher evaluation that leads to real changes in teacher behaviors. The study explored motivational forces and psychological processes related to the change process in adult behaviors. Data were collected by documenting the action research process and interviewing teachers at a suburban elementary school in Southern California. Findings suggest that through collaboratively developing a teacher evaluation model in a safe and supportive environment, teachers can enjoy the process and note real changes in behaviors as well as growth in student achievement. Key themes emerged, suggesting benefits associated with peer collaboration, goal-setting, evidence collection (portfolios), and ongoing and iterative process.
The dissertation of Nicholas Aron Bruski is approved.

Wellford (Buzz) Wilms

James Stigler

Noreen Webb

Christina Christie, Committee Chair

University of California, Los Angeles

2012
Table of Contents

ABSTRACT .................................................................................................................. ii

LIST OF FIGURES AND TABLES ............................................................................ ix

VITA ............................................................................................................................ x

CHAPTER 1 ................................................................................................................. 1

Introduction and Problem Statement ................................................................. 1
Purpose and Research Questions ................................................................. 2
The Research Site .......................................................................................... 3
Significance of the Problem .................................................................. 4
Research Methods ............................................................................... 6

CHAPTER 2: LITERATURE REVIEW ..................................................................... 10

Introduction ................................................................................................. 10
The Importance of Teacher Evaluation .......................................................... 10
Underlying Theory of Teacher Evaluation ....................................................... 12
  Conception of Teaching and Learning Must Inform the Research and Evaluation Design ........................................... 12
  Changing Teacher Behaviors .................................................................. 13
  The Purpose of Teacher Evaluation ......................................................... 16
  Systems and Feedback Loops .................................................................. 17
Research on Best Practices in Teacher Evaluation ........................................... 20
  Multiple Measures ................................................................................. 20
  Portfolios ............................................................................................... 20
  Video ....................................................................................................... 23
  Student Achievement/Value-Added Measurements .................................. 24
  Self Assessment and Reflection ................................................................ 28
  Goal Structures ...................................................................................... 29
  Peer Review ............................................................................................ 32
  Feedback .................................................................................................. 36
  Clearly Defining “Good Teaching” ............................................................ 38
Conclusion ....................................................................................................... 40

CHAPTER 3: METHODS ......................................................................................... 42

Introduction ....................................................................................................... 42
Research Questions ...................................................................................... 42
Research Design .......................................................................................... 43
Site Selection ................................................................................................. 45
CHAPTER 4: FINDINGS ................................................................. 61

Introduction .............................................................................. 61
Summary of Findings ............................................................... 63
Findings ...................................................................................... 64

RQ1: What are the Steps in the Process of Teachers Co-Creating a Teacher Evaluation System? .................................... 64
  Stage 1: Whole staff pre-survey and initial interviews .......... 64
  Stage 2: Literature review and journal responses ............... 65
  Stage 3: Group discussion and system design and review ...... 65
  Stage 4: Implementation of pilot via study groups .......... 68
  Stage 5: Mid-implementation study group check-in .......... 69
  Stage 6: Portfolio sharing and individual reflections .......... 69
  Stage 7: Group reflection and final interviews .............. 70

RQ2: What do Teachers at Mountainview Unified School Consider to be Important in Evaluating Teachers? ....................... 70
  Teachers do not believe that value-added measures of student achievement and teacher rankings or grade are useful evaluation tools .................................................. 71
  Participants felt current standardized assessments should play no part in teacher evaluation, but a test more focused on critical thinking and problem solving would be an improvement .................................................. 80
  Participants felt that even if standardized assessments measured what they view as important, the scores still should not account for most of a teacher’s evaluation .......... 83
  Teachers primarily view themselves as professionals .......... 84
  The primary purpose of an evaluation system should be teacher growth ................................................................. 86
  Systems of evaluation must develop teacher’s ownership of the process and motivate them to engage in thinking ................................................................. 88
  Teachers want to engage in this growing process in a safe and supportive environment ................................................................. 91
  A system of evaluation needs to be flexible in order to meet the varying needs of diverse teachers ................................................................. 98

RQ3: What do Teachers at Mountainview Unified School Believe are Appropriate Ways to Evaluate Teachers? ............... 101
  Teachers believe in peer observation and collaboration as appropriate and motivating facets of teacher evaluation ........ 102
A system of evaluation should be ongoing and frequent ........................................ 108
Reflection is an important part of learning and growing ..................................... 110
Teachers saw value in the use of portfolios as an evaluative device that provides clear evidence of teacher and student growth ................................................................. 111

RQ4: After Participating in an Action Research Model to Develop a Teacher Evaluation System, how Satisfied are Teachers with the Co-Developed Teacher Evaluation Process? ........................................................................................................ 114
Teachers were more than satisfied with the evaluation system ................................................................. 114

RQ5: What are the Challenges and Lessons Learned in the Process of Teachers Co-Creating a Teacher Evaluation System? ........................................................................................................ 117
Participants identified specific logistical components that they felt would lead to success as well as key elements of a successful evaluation system ................................................................. 117
Precision of language and clarity are highly important ........................................ 119
The process of reading the literature review changed and/or strengthened teachers’ beliefs about evaluation ........................................ 121
Participants were more likely to change their goals if they did not come with the forms already filled out ................................................................. 122
Participants felt the action research process led to increased buy-in to the co-developed teacher evaluation system ........................................................................................................ 124
The staff members were initially concerned about the time and resources that would be necessary to support a system of teacher evaluation, yet did not feel this process was a burden when reflecting afterwards ................................................................................................. 126
Efforts towards achieving evaluation goals infiltrated other aspects of teachers’ work ........................................................................................................ 127

CHAPTER 5: DISCUSSION .................................................................................................................. 128

Summary of the Findings .................................................................................................................. 128
Discussion ........................................................................................................................................ 130
Teacher Growth and Student Achievement ..................................................................................... 131
Creating Psychological Safety ........................................................................................................ 136
An evaluation system aligned with teachers’ conception of teaching can increase psychological safety ........................................................................................................ 137
Psychological safety can be enhanced when the administrator acts as a facilitator .......... 138
Co-creating a system of evaluation through an action research process creates increased psychological safety ........................................................................................................ 138
The flexible nature of the developed system led to psychological safety and satisfaction ........................................... 142
The Four Motivational Forces .......................................................... 143
The power of peer collaboration ...................................................... 144
Goal setting and evidence collection ............................................. 147
Frequency - an ongoing and iterative process ............................... 151
Learnings, Habits, and Thinking Transcended this
Process and Infiltrated other Areas of Participants
Teaching ...................................................................................... 154
Connections to Research Questions ............................................. 155
What do Teachers at Mountainview Unified School
Consider to be Important in Evaluating Teachers? ...................... 155
What do Teachers at MUS Believe are Appropriate
Ways to Evaluate Teachers? ....................................................... 156
After Participating in an Action Research Model to
Develop a Teacher Evaluation System, How Satisfied
are Teachers with the Co-Developed Teacher
Evaluation Process? .................................................................. 157
What are the Steps, Challenges, and Lessons Learned
in the Process of Teachers Co-Creating a Teacher
Evaluation System? .................................................................. 157
Steps ......................................................................................... 157
Challenges ................................................................................. 158
Lessons Learned ......................................................................... 159
Recommendations for Policy and Practice .................................... 165
Logistical Outline of an Evaluation System .................................. 165
Work Collaborative With Teacher to Explicitly Define
the Purpose of Evaluation ............................................................ 166
Involve Teachers in the Development of the
Evaluation System ....................................................................... 167
Teachers should Experience Evaluation through
Collaboration with Peers ............................................................. 167
Teachers Should Set Goals and Collect Multiple
Forms of Evidence to Demonstrate Progress
Towards those Goals .................................................................. 168
A System of Evaluation Should be Ongoing and Frequent ............ 169
The System Must be Flexible ....................................................... 170
Conclusion .................................................................................... 171
Limitations .................................................................................... 171
Lack of Definitive Impacts on Student Achievement ................. 172
Recommendations for Further Research .................................... 172
Dissemination of Findings ............................................................ 174
APPENDIX A: TIMELINE OF EVENTS .......................................... 175
APPENDIX B: WHOLE STAFF PRE-SURVEY PROTOCOL .............. 176
APPENDIX C: INTERVIEW #1 PROTOCOL ......................................................... 178
APPENDIX D: JOURNAL PROMPTS – LITERATURE REVIEW .......................... 180
APPENDIX E: GROUP DISCUSSION PROTOCOL ............................................. 181
APPENDIX F: MID-IMPLEMENTATION CHECK-IN JOURNAL PROMPT .............. 182
APPENDIX G: FINAL INTERVIEW ................................................................. 183
APPENDIX H: DRAFT 1 OF EVALUATION PILOT ........................................... 185
APPENDIX I: PROFESSIONAL GROWTH PLANS ......................................... 195
APPENDIX J: PRE-SURVEY RESPONSES ..................................................... 209
APPENDIX K: CODING CATEGORIES AND SUB CATEGORIES ......................... 213
REFERENCES ................................................................................................. 215
List of Figures and Tables

Table 3.1: Participant Demographic Data 59
Table 3.2: Data Collection Methods 60
Figure 3.1: Timeline of Events 63
Figure 4.1 Timeline of Events 69
Figure: 4.2 Student Writing Sample 85
Figure 5.1: Model of Teacher Growth 134
Figure 5.2: Checks and Balances to Teacher Growth and Student Achievement 135
Figure 5.3: Psychological Safety 140
Figure 5.4: Motivational Forces Leading to Components Required for Changes in Adults Behaviors 147
Vita

Nicholas A. Bruski attended Traverse City Senior High, Traverse City, MI. In 1996, he entered Michigan State University to study general business administration and pre-law. He later studied at the University of Michigan in the areas of social studies and English. He earned his teaching credential from the University of Phoenix in Los Angeles, CA in 2003 and later earned his Masters degree in Education from Walden University in 2008. He has served as a teacher, mathematics and data coach, principal, and is currently the Chief Academic Officer of a Southern California school district.
CHAPTER 1

Introduction and Problem Statement

A recent *LA Times* article on “value-added”¹ teacher evaluation has spurred a national debate on evaluating teachers. The *Times* published individual teacher “scores” ranking them based on the amount of growth of their students’ standardized test scores. Furthermore, the recent Documentary “Waiting for Superman” has received much attention, blaming teachers’ unions for much of today’s problems in public schools. Even as early as 1983, researchers were blaming collective bargaining for turning the evaluation process into a rule-based process of documenting minimum proficiencies rather than truly judging teachers and improving their competence (Darling-Hammond, Wise, & Pease, 1983).

In California, the Stull Act requires that probationary teachers be evaluated once a year while permanent teachers be evaluated every other year. However, there is great diversity in how teachers are evaluated as well as the criteria upon which they are evaluated. Many agree that current systems of teacher evaluation are ineffective and inefficient (Danielson, 2001; Hazi & Rucinski, 2009; Iwanicki, 2001; Marshall, 2005). There is also much disagreement on how teachers should be evaluated. Most states require evaluations every two to three years (Cohen, Walsh, & Biddle, 2008). Only 13 states require annual observations. Within these requirements, most states do not dictate how many evaluations occur during an evaluation year. In California, an administrator can observe a teacher for 20 minutes once during the year and still meet California Education Code. Thus, a teacher can be judged based on an observation of less than 0.1% of his or her

¹ “Value added” refers to documenting the growth students make in an academic year, typically measured by standardized test scores.
actual teaching. Although most states have policies on the frequency of evaluations, very few address the content of evaluations. To further illustrate the ineffectiveness of the teacher evaluation system in California, only 100 dismissal hearings occurred in California between 1996 and 2005 (Bathen, 1999). With over 300,000 teachers, less than 0.033% of teachers were found unfit to teach through this process in nearly a 10-year period.

In the Mountainview Unified School District, the evaluation process has fallen below the minimum benchmarks set by the Stull Act\(^2\). At a staff meeting in September of 2010, a teacher who had taught at the school for over 15 years commented, “I don’t think I’ve ever been evaluated.” A review of personnel files indicated that between 2002 and 2010 no teacher had more than two evaluation documents on file. These documents generally contained three to four sentences with ambiguous statements such as “The teacher differentiated instruction for multiple levels.”

**Purpose and Research Questions**

The purpose of my research was to develop a system of supervision and evaluation that meets the intents and statutory requirements of the Stull Act while incorporating research that supports an evaluation process that may lead to changing teacher behaviors to ultimately improve student learning. My research sought to address the following research questions:

1. What are the steps in the process of teachers co-creating a teacher evaluation system?

---

\(^2\) The Stull Act requires school districts to evaluate and assess the performance of certificated employees that perform the requirements of educational programs mandated by state or federal law (Education Code Sections 44660 to 44665). See Appendix A for complete details.
2. What do teachers at Mountainview Unified School consider to be important in evaluating teachers?

3. What do teachers at Mountainview Unified School believe are appropriate ways to evaluate teachers?

4. After participating in an action research model to develop a teacher evaluation system, how satisfied are teachers with the co-developed teacher evaluation process?

5. What are the challenges and lessons learned in the process of teachers co-creating a teacher evaluation system?

The Research Site

Darling-Hammond, Wise and Pease (1983) suggest that the success of any evaluation approach depends on the specific organizational context in which it is applied. The Mountainview Unified School District was poised for improvement in the realm of teacher evaluation. In 2009, the school hired consultants to examine its structures and processes in order to maximize effectiveness. After analyzing current practices and the results of parent and staff surveys, the board adopted the recommendations of the consultants and reorganized the administrative structure. The role of the school principal was split into two positions, a Dean and Chief Academic Officer. The Dean is responsible for day-to-day activities including facilities and student discipline. As Chief Academic Officer, I have the responsibility for teaching and learning, and working with teachers to improve their practice individually as well as through developing professional learning communities. The teachers are eager for someone to help them become the best teachers they can be. In addition, the school’s strategic plan calls for the development of a “revised
staff performance document that reflects the mission, vision and core beliefs of the district.” The coalescence of the national debate on teacher evaluation, the history of this particular school, and the new administrative structure provided a unique opportunity to develop and test an improved model of teacher evaluation.

**Significance of the Problem**

The current disconnected system of drop-in classroom visits and infrequent evaluation that equates classroom observations with evaluation leaves much room for improvement. The evaluation process needs to shift from a focus on quality assurance to recognizing, cultivating, and developing good teaching (Danielson, 2001). We must move from a system of detecting and preventing incompetencies to one of correcting them. Generally speaking, the system should shift from a summative to a formative approach. The infrequency of observation and feedback cycles makes the use of formative evaluation difficult. According to Iwanicki, “Teacher evaluations are most effective when they connect to student achievement and align with professional development and school improvement” (2001, p. 57). Marshall (2005) points out that in the current system, the principal or administrator owns the feedback and evaluation rarely changes teacher behavior. Even as far back as the early 1900’s, John Dewey (1910) claimed that learning must occur through engagement in real-world problem solving. Both students and adults must learn by applying their intelligence and efforts to dilemmas and complexities while reflecting on their experiences. Through this research, I will involve teachers in the creation of a process of supervision and evaluation that is focused on developing good teaching, empowering them to be involved in the process.
Many researchers conclude that teachers have an enormous impact on student achievement and there is great variability in teacher effectiveness (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008; Hattie, 2003; Hazi & Rucinski, 2009; Kane, Rockoff, & Staiger, 2007; Laczko-Kerr & Berliner, 2002; Wilson & Floden, 2003). The evaluation process can be one of the most effective and efficient ways for school leaders to assess teachers’ performance (Cohen, et al., 2008). This process should serve as an opportunity for growth and development that helps teachers improve their practice. Because Mountainview Unified School District ranks in the top decile of all California schools on standardized tests, some might ask “Why is this a problem?” As we would differentiate for our high achieving students, so must we differentiate for our high achieving teachers. Administrators have the responsibility to help teachers achieve their highest potential and constantly strive to improve their practice. It is also helpful to think about this question in terms of a value-added model of student achievement. The students at Montecito come to school with many more advantages than the typical student. Because they bring so much with them, there may be an opportunity to push them even farther than their already high levels of achievement.

In a loosely-coupled system such as public education, many barriers to organizational change exist. Weick (1976) describes education as a loosely coupled system. Educational institutions are made up of many parts that are responsive to each other, but also maintain their own identity. One method of overcoming the barriers to change in a loosely-coupled system is through the use of action research (Piggot-Irvine, 2010). I used action research in the development of an evaluation model at MUS. McNiff and Whitehead (2006) describe action research as “a form of enquiry that enables
practitioners everywhere to investigate and evaluate their work” (p. 7). It is a cycle of observing, reflecting, acting, and evaluating. It is a cycle of continuous growth and improvement.

An underlying assumption of any system of evaluation is that by providing feedback on strengths and weaknesses, the subject is expected to use the information to improve or change. In order to best develop a system of evaluation that leads to true change, it is also important to understand the work of Schein (1995) and his elaboration on Kurt Lewin’s (1952) change model. Lewin describes the change process of unfreezing learned behaviors, changing them, and then refreezing new behaviors through a psychological lens that considers the ego, thoughts, perceptions, feelings and attitudes. Psychological defenses, group norms and culture play an important role in the change process and will need to be considered when designing a system of teacher evaluation.

**Research Methods**

At MUS, I gathered a group of eight teachers to conduct the action research. We began by conducting initial interviews to collect data on teachers’ conceptions of teaching as it relates to themselves and the school as well as their experience with, and thoughts about, teacher evaluation to establish baseline data. We then studied the literature review together to develop a common understanding of research-based recommendations regarding teacher evaluation and supervision. Specifically, began by focusing on Darling-Hammond, Wise and Pease’s (1983) argument that a thorough understanding of context is critical in designing successful teacher evaluation. Common views of the teaching-learning process must be articulated in order to determine what the organization values and believes, which in turn helps determine how teachers are measured. I also exposed the
group briefly to the action-research process and its merits through an investigation of Piggett-Irvine’s (2010) use of action research in a New Zealand school to develop an improved appraisal system through a case study. Next, the group designed a model and instruments for evaluation and tested them. They then reconvened to evaluate the system of evaluation and further refine it in a cycle of continuous improvement. Throughout the cycle, I documented the process as well as surveyed the feelings and perceptions of the participants. Finally, I conducted final interviews revisiting themes from the initial interview to see how thoughts and beliefs had changed over time, as well as investigated participants feelings towards the process as a whole.

One specific organizational context at MUS that must be understood is the role of Chief Academic Officer acting as a coach and participant in professional learning communities. Martin and Taylor (2009) assert that coaches can be key facilitators in changing teacher behaviors by being responsive to the unique needs of specific sites and teachers. Consistently visiting classrooms and collecting data provide intimate knowledge of specific needs. Specific attention was given to what roles the Chief Academic Officers plays in the evaluation and supervision process. Although research supports the benefits of coaching teachers to increased effectiveness, others warn about the inherent conflict between coaching and evaluating (Danielson, 2001; Hazi & Rucinski, 2009).

Like Darling-Hammond, Wise and Pease (1983), Martin and Taylor (2009) speak to the necessity of understanding teachers capacities and belief systems as well as the overall context in which teaching and learning take place. However, they specifically advocate for the use of data in establishing professional development needs. According to the authors,
examining data and evidence to collaboratively identify growth needs is key to sustaining change.

More frequent observations and building accountability through goal setting and data can create tensions in supervision and evaluation system. Fenwick (2001) identified three specific tensions in a Canadian province that mandated teacher professional growth plans (TPGP’s). Generally, all three tensions focus on the balance between teacher autonomy and control and administrative regulation, oversight, and direction-setting. However, the teachers and administrators in this study found ways to balance these tensions, with the majority reporting that although they felt anxious or ambivalent in the early stages, they felt empowered and enthusiastic after three years of implementation.

In order to ensure an improved system of evaluation at Montecito, it was critical to review research in the area of changing teacher behavior through supervision and evaluation. The review of literature helped shape the direction of the proposed research at MUS. Piggot and Irvine’s (2010) success in using action-research to develop a system of supervision and evaluation justified the use of action-research here. With an experienced and professional staff, it was critical to include them in the development of the system to create buy-in and ownership. However, Fenwick’s (2001) experiences with guiding goal-setting to align with district goals had implications in this work. MUS was and is in the process of trying to make the newly created strategic plan a living and breathing document. This process had to be thoughtfully structured so that the work supported the initiatives in the strategic plan while providing teachers with the autonomy to have authentic input into the process. I also wanted to ensure that the decisions made and structures developed are
research-based. Sharing and studying the literature review together to identify successful practices while tailoring a plan to the unique needs of MUS helped accomplish this.

This work will culminate with an adoption of a system of evaluation at Mountainview Unified School. I will then share the research at a Santa Barbara County Curriculum Council meeting so that others in the area will benefit from my research. I also propose to submit a summary to scholarly journals for publication. The value of this particular model is that it could potentially be used in a wide variety of educational settings with diverse demographics. The same process may yield different, yet effective, models of teacher evaluation in other districts.
CHAPTER 2

Introduction

In this literature review, I provide an analysis of extensive research showing the necessity of improving the current system of teacher evaluation. By exploring Darling Hammond’s (1983) work on defining perceptions of teaching, I demonstrate why having a clear understanding of Mountainview Unified’s perception of teaching is important in informing evaluation practices. I also discuss how Schein’s (1995) expansion of Kurt Lewin’s change model guides the interpretation of the current research in the field of teacher evaluation. Finally, I explore how these theories relate to the research on best practices in teacher evaluation in areas of peer review, using student achievement data and value-added systems, portfolios, videotaping lessons, feedback and promoting self-reflection and assessment.

The Importance of Teacher Evaluation Research

Teacher supervision and evaluation is a topic that has received consistent attention in education journals in recent years. Throughout these articles, there is unanimous agreement that current practices are ineffective (Britton & Anderson, 2009; Fenwick, 2001; Kohut, Burnap, & Yon, 2007; Marshall, 2005; Weems & Rogers, 2010). In California, the Stull Act minimally requires:

- Evaluations once each year for probationary certificated employees
- Every other year for permanent certificated employees
- Every five years for certificated employees with permanent status who have been employed at least ten years with the school district, are highly qualified (as defined in 20 U.S.C. § 7801), and whose previous evaluation rated the employee as meeting or
exceeding standards, if the evaluator and certificated employee being evaluated agree.

Typically, tenured teachers are observed by administrators once a year to meet the State’s evaluation requirements. An administrator observing one 20-minute lesson in a year would equate to observing 0.1% of a teacher’s instruction (Marshall, 2005).

The current prescriptive system of evaluation focuses on documenting negative findings leads to inflexibility and anxiety (Darling-Hammond, et al., 1983). The authors assert that we must move from a system of detecting and preventing incompetencies, to one of correcting them. Generally speaking, the system should shift from a summative to a formative approach. The infrequency of observation and feedback cycles makes the use of formative evaluation difficult. Secretary of Education Arne Duncan summarizes, saying, “Our system of teacher evaluation frustrates teachers who feel their good work goes unrecognized and ignores teacher who would benefit from additional support” (The New Teacher Project, 2010, p. 3). Ineffective systems of teacher evaluation are negligent in their wasted opportunity to positively impact teacher effectiveness and student learning. Researchers unanimously agree that the classroom teacher is the single most powerful factor in student achievement that we have control over (Darling-Hammond, 2010; Hattie, 2003; The New Teacher Project, 2010). We owe it to our students to do everything we can to promote teacher growth and learning. An effective system of teacher evaluation is a prime vehicle for creating conditions to support teachers in strengthening their pedagogy.
Underlying Theory of Teacher Evaluation

Conception of Teaching and Learning Must Inform the Research and Evaluation Design

In order to establish an effective system of evaluation at MUS, it is critical to understand what teachers believe about the purpose of teaching as well as the purpose of evaluation. By first understanding their beliefs on teaching and evaluation, instruments and procedures can be aligned to their framework of teaching. This study will employ a qualitative action research design to work with teachers, co-constructing a research-based system of teacher evaluation. The use of action research will allow me to document the teachers’ beliefs about teacher evaluation as well as their perceptions of the process, lending itself to constructivist or participatory knowledge claims. Action research is an appropriate approach in that the learning needs and direct feedback from participants will guide and direct the content and structure of the research. McNiff and Whitehead (2006) describe action research as “a form of enquiry that enables practitioners everywhere to investigate and evaluate their work” (p. 7). It is a cycle of observing, reflecting, acting, and evaluating.

Darling-Hammond, Wise, and Pease’s comprehensive Teacher Evaluation in the Organizational Context: A Review of Literature (1983) proves to be a seminal study advocating for alignment between educational theory and educational processes by understanding the organizational context. They argue that a thorough understanding of context (i.e. teachers’ beliefs about the nature of teaching) is critical in designing successful teacher evaluation.
Common views of the teaching-learning process must be articulated in order to determine what the organization values and believes, which in turn helps determine how teachers are measured. In designing a system of evaluation, an organization must understand how it conceptualizes the work of teaching. Under the conception of teaching as labor, teaching activities are routinized and prescribed. The evaluation system acts as a direct inspection of the work (Darling-Hammond, et al., 1983). When teaching is conceptualized as craft, teachers are expected to have techniques with generalized rules for application. In this system, evaluation is used to ensure that teachers have these prerequisite skills. In teaching as a profession, teachers are expected to have these prerequisite skills, but also to exercise judgment in their application. Teachers need theoretical knowledge as well as practical techniques. The evaluations process becomes more of a judgment on the ability of teachers to problem-solve. In teaching as art, teaching techniques and the ways they are applied may be novel and unconventional. There is a shift away from standardization where teachers have to incorporate their pedagogical knowledge and their own personal resources and intuition. In this conceptual framework, evaluation involves self-assessment and critical assessment by multiple people.

In reality, these ideal types of conceptions of teaching will not be found in pure form, but rather as mixes, sometimes depending on which job functions a teacher is fulfilling. When being evaluated on lesson creation and revision, a school may see their work as art, but when being evaluated on accuracy of attendance procedures, completing paperwork, and other routines, this may be seen as labor. Understanding which of the four conceptions of teaching Montecito teachers identify with and when will be critical in successfully designing a system of evaluation.
Changing Teacher Behaviors

An underlying assumption of any system of evaluation is that by providing feedback on strengths and weaknesses, the subject is expected to use the information to improve, or change. In order to best develop a system of evaluation that leads to true change, it is important to understand the work of Schein (1995) and his elaboration on Kurt Lewin’s (1952) change model. Lewin describes the change process of unfreezing learned behaviors, changing them, and then refreezing new behaviors through a psychological lens that considers the ego, thoughts, perceptions, feelings and attitudes. Psychological defenses, group norms and culture play an important role in the change process.

Schein (1995) adds to this body of knowledge by describing each stage in more depth. Change begins with disconfirmation, or dissatisfaction and frustration that arises from data that does not meet one’s expectations and hopes. In the realm of teaching, this could include feedback given to the teachers on student achievement or observable behaviors in the classroom either by teachers or students. Dissatisfaction with one’s own teaching might also be induced by observing other teachers who one sees as superior to his or her own teaching. However, disconfirmation alone is not enough to change behavior. Some may ignore the information or blame it on external factors. In order to accept the disconfirming evidence, it must connect to something that the teacher cares about, inducing a feeling that not changing will lead to failure of some sort. Schein refers to this feeling as survival anxiety. By accepting our failures, we are in danger of losing effectiveness and losing face with our peers. However, adapting poorly can sometimes look preferable to risking failure or losing self-esteem by trying and failing in new endeavors. In
a system of teacher evaluation, we want teachers to take risks and stretch themselves beyond their comfort zones.

To combat these maladaptive behaviors, efforts need to be made to create *psychological safety* (Schein, 1995). The threat of the disconfirming data must be balanced with appropriate psychological safety so that a teacher can accept the data, feel survival anxiety, and thus become motivated to change. Psychological safety can be created in many ways. Working in groups leads to increased psychological safety. Peer review and collaboration within an evaluation system might be one way to support changing teacher behaviors. Schein also advocates for creating spaces where errors and risk-taking are embraced and encouraged. Designing a system and culture that communicates this is critical to supporting changes in behavior. Providing on-going support, or coaching, can also reduce learning anxiety and create the safety teachers need to take risks.

Once teachers have accepted the data and feel an appropriate balance of anxiety and safety, they broaden their knowledge through *cognitive redefinition* (Schein, 1995). This includes learning new meanings of terms, broadened interpretations of concepts, and new anchors and scales of judging and evaluating. In the teaching realm, this might include new concepts of what “understanding” means to students or new benchmarks for what constitutes satisfactory learning and behavior in a classroom. One way in which these new ways of learning are introduced is through imitation and identification with a role model. In an evaluation system, the use of peer review and support is one way to introduce new learning in this way. When peers engage in meaningful conversations and share their interpretations of lessons and teaching activities with one another, they can be challenged to imitate positive role models.
Another means of learning that can occur when role models are not present is *scanning* (Schein, 1995). Scanning occurs by reading articles and books, attending conferences, or going back to school. Teachers can use their new learning in trial and error to further develop and practice. This trial and error can reinforce new practices or lead to a new cycle of disconfirmation.

**The Purpose of Evaluation**

When designing a system of supervision and evaluation, it is important to understand its purpose. One difficulty with the current system is the dichotomous nature of ensuring teaching quality while simultaneously trying to promote professional development (Danielson, 2001; Weems & Rogers, 2010). The pressure of performing for evaluation can inhibit risk-taking needed to foster professional growth (Brogan, 1995). Although professional development and accountability are implied, other researchers put more emphasis on student achievement (Feeney, 2007; Iwanicki, 2001). However, by focusing on providing teachers with opportunities to learn, all of these purposes can be met (Darling-Hammond, 2010). When evaluation is structured to help teachers learn, they will grow professionally and student achievement will increase. Knowing the purpose of evaluation should inform the structure of the evaluation system. Weems (2010) states that if the sole purpose of evaluation is identifying competence, then observations by principals is adequate. However, to move beyond simply checking for competence, more diverse methods of evaluation are needed.

A second subtle nuance in teacher evaluation is the importance of focusing on observable measures of actual classroom instruction versus the thinking, planning, and reflecting that happens before and after the actual teaching. Marshall (2005) believes that
evaluation should foster openness to feedback, develop a “supervisory voice” on reflection in teachers’ heads, and instill an hyper-awareness of student learning results. Besides pointing out successes and challenges, evaluation should help train teachers to understand what they need to do to improve their practice (Rebora, 2010). An evaluation system that focuses on developing processes for thinking and reflection will better serve teachers by giving them skills that they can apply to every facet of teaching. “We need to create conditions, even inside large organizations, that make it possible for individuals to get the power to experiment, to create, to develop, to test – to innovate” (Moss Kanter, 1983, p. 15). The traditional method of observing a lesson and providing feedback on one specific lesson may serve to improve that particular lesson, but often is not enough to change the way a teacher teaches, or thinks about teaching.

**Systems and Feedback Loops**

In addition to understanding educators’ beliefs about teaching and evaluation, understanding theory behind feedback and motivation is also important to consider when designing a system of supervision and evaluation. Classroom teaching can be thought of as a system. Kauffman (1980) defines a system as “a collection of parts which interact with each other to function as whole,” (p. 1). A system is made up of inputs and outputs. Information about the outputs of a system is fed back into the system, which creates a continuous cycle of learning. In this way, a system can be responsive to the environment by analyzing information provided by feedback and adjusting as necessary.

Our current teacher evaluation system is one such system that consists of inputs, outputs, and feedback. One weakness of the system of teacher evaluation is the length of time between feedback cycles. With observations occurring only once every one to five
years, this minimal feedback is unlikely to truly change behavior. Furthermore, the structure and language of the evaluation forms emphasizes single-loop learning, where the evaluator scans the environment (observes), compares the information against operating norms (teaching standards and accepted practices), and initiates appropriate action (grades on evaluation forms) (Kauffman, 1980). This system does not encourage teachers and principals to question whether or not the operating norms, or standards, are appropriate. It simply measures the observation notes against a predetermined list of norms.

One failure of this system of evaluation is that it operates on the assumptions of MacGregor’s Theory X (Northouse, 2007). Management assumes teachers are inherently lazy and dislike work. Problems are viewed as deficiencies in people, not the system. Therefore, a hierarchical structure is needed to closely supervise them. The administrator is the only person deemed qualified to evaluate their work and seeks to reward or punish them through their written evaluation. In contrast, Theory Y believes that employees enjoy work and can creatively problem solve. For Theory Y to work in an organization, the hierarchy between administrator and teacher needs to be flattened, allowing for trust and cooperation to occur between the two.

Fairtlough (2007) suggests an alternative to the traditional hierarchical structure that embraces the tenets of Theory Y. In a responsible autonomy, individuals or groups have autonomy to decide what to do, but are accountable for their decisions. Some may believe that education in general is a responsible autonomy in which teachers choose how to educate their students. On the spectrum of hierarchy vs. autonomy, education appears to be sliding closer to hierarchy through mandated curriculum, state standards, high-stakes
testing and other grant-funded initiatives where participation is tied to strict compliance with a set of policies. However, it is possible to work within the confines of the hierarchical system and still promote responsible autonomy.

One element of responsible autonomy that Fairtlough (2007) describes is critique. When an external agency evaluates the autonomous unit, it encourages that unit to act responsibility by holding it accountable for its actions. This structure currently exists in my school district. I am the external agent who evaluates. However, Fairtlough warns that too lax a critique allows some to be overlooked with an inaccurate picture of performance. This occurs with the infrequency of formal observations. Through anecdotal and personal experience, I have found that teachers will put on their best show when being formally observed. Lesson plans are thoroughly written, extensive materials are prepared, and students are even coached on how to behave during the observation. Increasing the frequency of observations could increase the accuracy of critique, and therefore the effectiveness of the responsible autonomy structure.

Providing teachers with time to collaborate, observe each other, and reflect more informally could also provide teachers with more opportunities for feedback. This would enable double-loop learning, where teachers could question practices as a whole, not just measuring themselves against an established norm (Kauffman, 1980). Although the process is constrained by the mechanics outlined in education code, collective bargaining agreements, and board policy, the actions that take place around these structures could be more focused on teachers and less hierarchical. By providing teachers with time to work collaboratively, a Complex Evolving System (Fairtlough, 2007) would develop through action research, where teachers could adapt and evolve their practice to create coherence.
This system would be much more responsive to the environment, adapting to the needs of current students and encouraging teachers to constantly reflect and improve their practice.

**Research on Best Practices in Teacher Evaluation**

**Multiple Measures**

Researchers agree that teachers must be evaluated on multiple measures (Sawchuck, 2009b; Schochet & Chiang, 2010; The New Teacher Project, 2010; Tyler, 2010; Weems & Rogers, 2010). Single observations or student test scores (such as value-added models discussed later) do not provide an accurate picture of teacher quality and effectiveness. If Mountainview Unified teachers identify with the concept of teaching as a *profession*, the evaluation system should be designed to measure not only their technical knowledge of teaching, but their theoretical knowledge and ability to problem solve and apply their knowledge to unique situation (Darling-Hammond, et al., 1983). In teaching as *art*, teachers incorporate personal resources and use intuition to create nuanced and individualized instruction. To evaluate teaching as *art*, other methods, such as self-reflection and writing are needed to access the thinking behind decisions.

**Portfolios**

The use of portfolios\(^3\) as a means of evaluating teaching through multiple measures is one method that is gaining traction in school districts, higher education and professional organizations as well as research (Brogan, 1995; Danielson, 2001; Darling-Hammond, 2010; Ediger, 2000; Feeney, 2007; Iwanicki, 2001; Milner, 2010; Rebora, 2010; Weems & Rogers, 2010). This might include samples of student work and test scores, lesson plans, units of study, videotape, and self-reflection writing. A strength of using portfolios to

---

\(^3\) For the purpose of this research, I define portfolios as a collection of evidence demonstrating teacher and student learning.
evaluate teachers is that it embraces a philosophy of contextualism, or giving meaning to observable facts (Ediger, 2000). Although little empirical research exists on the effects of using teacher portfolios, it is a technique that has been embraced by leading researchers in education as well as National Board Certification, the highest honor in teacher credentialing nationally. Furthermore, even though portfolios as a whole have not received much attention in research, individual components of portfolios such as using video, self-reflection, student data, and peer review have. All of these factors merit consideration of the use of portfolios as a strategy in teacher evaluation at MUS.

One strength of using portfolios is that the process of collecting evidence and defending its choice forces teachers to reflect on their practice (Danielson, 2001; Darling-Hammond, 2010; Weems & Rogers, 2010). Iwanicki (2001) urges portfolios to include evidence of a teacher’s intents, processes, and outcomes in teaching a lesson, unit, or even an academic year. This aligns closely with teaching as a profession and teaching as art; teachers are required to show evidence of problem solving, how they used intuition in designing their teaching, and what they have learned from those experiences. When teachers are asked to write about why they designed a lesson or unit in a particular way and how it could be improved, they are forced to reflect on their practice. This activity can serve as the disconfirmation needed to begin the change process by illuminating opportunities for growth.

Darling-Hammond (2010) outlines a five-step portfolio outline of what might be included in a portfolio. This includes thinking about the context for learning, planning instruction and assessment, instructing students and supporting learning, assessing student learning, and reflecting on teaching. Artifacts as evidence might include teacher
commentary throughout all five steps, lesson plans and instructional materials, video clips, student work samples, evaluative criteria or rubrics, and daily reflections. Brogan (1995) advocates for including a goals statement, philosophy of teaching, summary of professional responsibilities, service to the school and profession, and criteria for assessing teaching performance, such as student work, assessment data, lesson plans, observation summaries, and more. Requirements can also be added to include connections to school or district initiatives and priorities. A portfolio requirement of evidence of advancing the work of the Mountainview Unified strategic plan could help to further the focus of our work on agreed upon priorities.

In addition to encouraging reflection, the use of portfolios requiring evidence of student learning should also help create a focus on outcomes, not just intentions. Any time a new strategy is implemented, evidence should be collected to assess its effectiveness (Feeney, 2007). Through the perspective of teaching as labor, and teaching as craft, a checklist of teacher behaviors and skills may be enough to evaluate effective teaching. However, simply using a prescribed technique does not ensure that the student learned the material. By asking a teacher to document evidence of learning, they are forced to ask themselves “How do I know if they have mastered this skill?”

Other tangential benefits claimed by proponents of portfolios include increased peer learning through evaluation of each other’s evidence as well as an increased sense of collaboration (Danielson, 2001; Darling-Hammond, 2010). The act of scoring a portfolio can lead to learning in both the scorer and the creator of the portfolio. The increased interaction between peers can also expose them each other’s strengths and weaknesses. This knowledge can lead to more effective interaction and learning from role models.
Although the defense of portfolios included above is more theoretical than empirical, a discussion of research on individual components of portfolio items can provide evidence of the utility of portfolio use in evaluating teachers.

**Video**

In schools with limited opportunities for peer observation, the use of a video camera to record lessons is a viable option for promoting peer review as well as self-assessment and reflection. Schools that do not have the structure or funds to relieve teachers from duties to observe each other during the instructional day can purchase video cameras for as little as two hundred dollars, effectively giving them access to each other’s classrooms. Although the use of video to evaluate teaching has been around since as early as 1975, it is a strategy that is underutilized in today’s systems of evaluation (Capizzi, Wehby, & Sandmel, 2010). The use of video can be a particularly effective means of providing disconfirming data to teachers, priming them for the change process. Teachers are often unable to objectively see themselves teaching and what is happening around the room while they are engaged in the act of teaching. Using video allows them to see themselves from another perspective and allows them to take the time to stop at any point in the lesson to think about their decisions, ask questions, reflect, and seek feedback from peers.

In a study of three graduate students specializing in special education, the use of video was found to increase the inclusion of critical lesson components and the process was rated with high levels of satisfaction (Capizzi, et al., 2010). The practice of using videos of teaching has also been embraced by the National Board Certification process, and is recommended by respected industry experts (Darling-Hammond, 2010; Rebora, 2010; Sawchuck, 2009a, 2009c). Current research funded through the Bill and Melinda Gates
foundation is underway to videotape 4,000 teachers in New York City, Charlotte, and other metropolitan areas, using the videos to identify best practices in teaching (Sawchuck, 2009a). Marshall states, “There is no better way to see the flaws in one’s teaching (and appreciate the strengths) than to watch a videotape with a critical friend” (2005, p. 734).

**Student Achievement/Value-Added Measurements**

An area receiving much attention in the media and research is the use of value-added models of teacher evaluation where teachers are graded on how far students progress in the academic year, typically measured by high-stakes state tests. The New Teacher Project, founded by Washington D.C.’s former chancellor of public schools, Michelle Rhee, published a report entitled *Teacher Evaluation 2.0* (2010) in which the group strongly advocates for the evaluating teachers based on the academic growth of students. Rhee met some controversy with her efforts to restructure teacher tenure, compensation, and evaluation in the capitol city. In 2010, 241 D.C. teachers were fired, most of whom had received poor evaluations. When the *LA Times* published teacher names and value-added scores, outrage was voiced by many, including teachers unions and the popular press. One teacher allegedly took his own life because of the embarrassment caused by the publication of his poor scores.

The logic behind measuring what value individual teachers add to their students’ knowledge base appears rational at face value. Tennessee has been using the Tennessee Value-Added Assessment System (TVASS) since 1993 to measure teacher contributions to student learning. A study of the TVAAS using mixed methods statistical longitudinal multivariate analysis found that teacher effects are dominant over classroom context variables such as class size and heterogeneity of student abilities. The study reported that
teachers are the most important factor affecting student learning and there is wide variation in teacher effectiveness. Throughout the research, there is wide agreement that teachers have an enormous impact on student achievement and there is great variability in teacher effectiveness (Boyd, et al., 2008; Hazi & Rucinski, 2009; Kane, et al., 2007; Laczko-Kerr & Berliner, 2002; Wilson & Floden, 2003). Knowing this, why is there controversy over using value-added measures in evaluating teachers?

One concern that many have for using value-added assessments in evaluation is the validity and reliability of the results (Darling-Hammond, 2010; Fuhrman, 2010; Hanushek & Rivken, 2010; Marshall, 2005, 2008; Milner, 2010; Schochet & Chiang, 2010). The Institute of Education Sciences conducted a hypothetical analysis of error rates in the use of value-added measures of teacher effectiveness. Using one year of data, type I and II error rates were approximated to occur at a rate of 35%. Additional years of data reduced these percents to near 20%. This means that between one fourth and one third of teachers with average performance will be erroneously identified as needing support. The Institute also synthesized the results of five studies on value-added measure and summarized that one half to one third of teachers in the top quartile or quintile one year would drop to the next lowest level the following year. This summary of research speaks loudly to the unreliability of using these measures for high-stakes decisions. When scores are used for decisions such as layoffs, promotions, and bonuses, other factors need to be taken into consideration.

Such errors in validity and reliability can occur for many reasons, including random differences among students such as background and ability, idiosyncrasies of individual classrooms such as barking dogs on test days or a particularly disruptive student over the
course of a year (Schochet & Chiang, 2010). Another complicating factor is that students typically are not randomly assigned to classes (Fuhrman, 2010). Placement decisions are made based on student and teacher differences. For example, a teacher with a reputation for impacting young boys without a father figure might get higher percentages of these students on an annual basis. To further complicate matters, imagine that the school has a counseling program specifically focused on the needs of boys like this in this particular grade. The teacher would potentially have an advantage when being evaluated on value-added measures as his students typically have opportunity for growth and additional support that others do not. Some populations are also more transient than others. Schools that serve military populations often have a revolving door of students throughout the year. It would be unfair to compare them to their peers at another school two miles away.

It is clear that the many omitted variables could never be completely accounted for (Hanushek & Rivken, 2010).

Another difficulty in using annual state tests for value-added purposes is that the results are typically not available until the summer or fall after the school year has ended. The delay of this feedback makes it difficult for teachers to use the information in any meaningful sense for improvement. Furthermore, in California, the tests are not taken until second grade and they are only given in language arts and mathematics, with the exception of adding science in fifth grade. Kindergarten and first grade teachers as well as specialists in music, art, physical education and other areas do not have this data to be evaluated on. Finally when high stakes are placed on specific tests, there is motivation to teach to the test (Hanushek & Rivken, 2010; Marshall, 2008).
Although it is clear that using high-stakes state tests to determine value-added measurements is fraught with complications, it does not mean that we should not focus on how teachers impact student learning. Using data to help teachers improve is a valid and recommended strategy (Britton & Anderson, 2009). Some even believe that teachers should establish what data should be collected (Marshall, 2005; Williamson & Blackburn, 2009). Marshall advocates for using before-and-after assessments, establishing a baseline and documenting growth. By asking teachers to submit evidence of student learning, the values of teaching as a profession and teaching as art are honored. Rather than using hard cut points and averages, teachers can use portfolios to share a story of student learning, incorporating student work from the becoming of a unit to work at the end of the unit. This could include class averages on multiple-choice tests or accuracy rates on timed tests, but it could also include student writing, projects, and other performances of understanding. Doing so can also promote collegiality by sharing successes and convincing peers that learning took place through peer evaluation. The competitive nature of high-stakes value-added models can lead to seclusion and an unwillingness to share when one is constantly worried that his or her neighbor might get a higher score (Marshall, 2008). Allowing teachers to share stories of learning through portfolios encourages collaboration while increasing the psychological safety necessary to learn and change. Martin and Taylor (2009) assert that consistently visiting classrooms and collecting data provides intimate knowledge of specific needs. Like Darling-Hammond, Wise and Pease (1983), Martin and Taylor speak to the necessity of understanding teachers capacities and belief systems as well as the overall context in which teaching and learning take place. However, they specifically advocate for the use of data in establishing professional development needs.
According to the authors, examining data and evidence to collaboratively identify growth needs is key to sustaining change.

**Self-Assessment and Reflection**

An implied but critical piece of portfolios, video, using student achievement data, and goal setting is using self-assessment and reflection as a tool for learning. Although reflection as a research paradigm has received increased attention, it has not successfully entered the practice of teaching (Siens & Ebmeier, 1996). Self-assessment is a powerful tool in producing disconfirmation, a critical first step in the learning process. Teachers must identify areas of potential growth, while also recognizing their strengths. Traditional evaluation systems often put teachers in a passive role of receiving information from a supervisor. Feeney states evaluations have no meaning unless there is reflection and, “Teachers will not be able to increase their capacity for accurate self-assessment in the wake of isolated external value judgments,” (2007, p. 193). In a study of the Chicago Excellence in Teaching Project, researchers found that self-assessment promotes learning (Danielson, 2010). This process of reflection converts the traditionally passive system of teacher evaluation to an active system of learning and growth.

Reflection is a skill that can be taught. In a study of 120 Midwest teachers, graduate student administrators received 30 hours of training in peer coaching. Phone interviews were coded and analyzed, giving each teacher a reflection score. The study found a statistically significant correlation between developmental supervision and reflection. The structure of knowing that they would have to talk about a lesson forced teachers to reflect on their practice in order to be prepared for questioning. In this study, teachers also
reported reduced feelings of isolation and had increased opportunities for feedback by regularly interacting with supervisors and peers.

When teachers constantly rely on others to evaluate them, their confidence to evaluate themselves can erode (Foran & Doylend, 2004). When conceptualizing teaching as a profession or teaching as art, including an element of self-assessment and reflection is a prerequisite. Both focus on the use of judgment and intuition that are not easily accessible to outside observers. Self-assessment and reflection are necessary tools to expose teachers’ thinking to outside observers.

**Goal Structures**

One technique that can help promote self-assessment and learning is the use of goals. Setting goals can help teachers begin with the end in mind and can provide structure to the use of portfolios (Brogan, 1995). Goals can be thought of as having two different structures (Wolters & Daugherty, 2007). Mastery goals are focused on the more technical components at the center of teaching as labor or craft. Performance goals measure practices, policies and procedures and can be used to access the nuances of teaching as a profession and art. Using a web-based self-report system of 1,024 teachers, Wolters and Daugherty found that the use of goals increased teachers’ sense of efficacy.

A study of three school sites in Alberta, Canada took a more structured approach to the use of goals in what they referred to as professional growth plans (Fenwick, 2001). Teachers were provided with an orientation on goal setting, were given faculty meeting time to develop and write growth plans in teams, and participated in mid-year “check-ins” on the progress towards their goals. Administrators were careful to avoid the perception that they were trying to shape teachers goals, but Fenwick still found that the goal setting
process had a normalizing effect. Sample goals shared at an orientation or the types of questions asked about goals can have the effect of leading teachers to believe that administration wants a certain type of goal. Goals were also written in teams and the co-development of goals with peers may have contributed to normalizing effects. These effects could be leveraged when trying to advance specific school priorities, such as those outlined in Mountainview Unified’s strategic plan.

Fenwick (2001) found that the use of professional growth plans developed a greater sense of commitment and focus among teachers. They felt that they were held more accountable, but also felt more collegiality. Many reported that trust was developed when administrators accepted their proposed goals and evaluation was shifted from a sense of surveillance and judging to one of praising and acknowledging growth. This technique could prove useful in developing the psychological safety necessary for learning. Although teachers reported anxiety in the initial stages of this process, high levels of comfort were reported after three years of use. Most teachers reported that the process built their confidence in their teaching abilities. However, some were concerned about the rigidity of the process and reported an unwillingness to take risks. To mitigate this effect, Schein (1995) recommends creating a culture where errors and risk-taking are embraced and encouraged. The plans also did not adjust for the unpredictable nature of learning. For example, if a teacher was inspired by a workshop or conversation, their new passions might not match the pre-written growth plan. This process also required a significant investment of time that might restrict others from adopting this model.

More frequent observations and building accountability through goal setting and data can create tensions in supervision and evaluation system. Fenwick (2001) identified
three specific tensions in a Canadian province that mandated teacher professional growth plans (TPGP’s). Teachers were required to create plans consisting of three professional development goals to help them become better teachers. Administrators acted as coaches, conferencing with the teachers throughout the year and informally visiting classrooms. First, allowing teachers to set their own goals promoted self-direction and autonomy, yet the increased surveillance implied more regulation of teaching and learning. Second, the idea behind the plans seemed to honor individual teachers’ unique ways of knowing while simultaneously reinforcing the idea that there are specific skills to be mastered in becoming a great teacher, which would be outlined through measurable objectives. This is similar the Darling Hammond, Wise and Pease’ teaching as labor or teaching as craft model (1983). Finally, the TPGP’s encouraged teacher risk-taking, creativity, and personally meaningful learning, but the policy also requires plans to align with school and teaching goals as well as professional teaching standards.

Generally, all three tensions focus on the balance between teacher autonomy and control and administrative regulation, oversight, and direction-setting. However, the teachers and administrators in this study found ways to balance these tensions, with the majority reporting that although they felt anxious or ambivalent in the early stages, they felt empowered and enthusiastic after three years of implementation. Administrators identified strategies that they believed help balance these tensions. They intentionally held back criticism of teacher-created goals focused on celebrating successes and giving gentle guidance. Through surveillance, principals also focused more on praising and illuminating progress than judging. When working to align teacher goals with district goals, administrators modeled their own plans as well asking teachers to consider how their
goals might align with district goals or strategic plans. However, they were conscious to develop a trusting relationship and ensure that the TPGP’s were authentic and meaningful records of teacher learning.

One challenge of setting goals early in the year is that teachers might not have the disconfirming data needed to motivate them to change or learn. By providing opportunities for peer review and observation (discussed below) prior to goal writing, teachers will have more disconfirming information with which to drive their goals.

**Peer Review**

Another strategy mentioned earlier that helps create psychological safety to support the change process is collaborative work (Schein, 1995). In teacher evaluation, the use of peer review can meet this need. Schein’s model of changing teacher behavior also discusses how learning occurs through imitation of role models. In the typical isolated classroom setting, there is little to no exposure to role models. A process of peer evaluation would provide students with access to role models, assisting with the process of cognitive redefinition. Arne Duncan, the US Secretary of Education said, “I’m a big fan of peer evaluation” (Sawchuck, 2009c, p. 7). However, due to financial and time constraints, peer review is a process that is infrequently used at the elementary level. Mountainview Unified is uniquely poised to utilize this strategy. Each teacher sends her class to a variety of specialists each week including music, art, physical education, technology, science and Spanish for periods of 30 to 60 minutes. This structure would allow the staff to observe each other on a regular basis without any additional cost to the school.

Because peer review is rarely employed at the elementary level, much of the research has taken place in alternate settings such as higher education and with pre-
service teachers where universities require or promote the practice. The research also generally focuses on participants’ perceptions of the process through qualitative surveys, interviews, and focus groups. However, the research needs to be scrutinized for generalizability.

The benefits of peer review suggested in research are many. When teachers, professors, or students are observed by their peers, it lowers the anxiety relative to being evaluated and observed by an authority figure (Britton & Anderson, 2009; Iwanicki, 2001; Kohut, et al., 2007; Marshall, 2005; Wilkins, Shin, & Ainsworth, 2009). With less anxiety, the process of evaluation serves to focus more on teacher development and collaboration with peers than traditional models that often serve as ex post facto summaries of 30-minute observations by administrators (Britton & Anderson, 2009; Byrne, Brown, & Challen, 2010; Hattie, 2003; Marshall, 2005; Sawchuck, 2009c; Wilkins, et al., 2009). However, not all peer review models are effective. In a study of peer review in the United Kingdom, Byrne (2010) survey 36 faculty members and found that most were dissatisfied with the current system of peer review, noting that it was perceived as a check box to be marked off to meet the demands of administration as well as the lack of valuable feedback peers were able to offer and an inflexible system. One important change that leads to increased satisfaction in this study was providing training to peers to become better at observing and providing feedback. Similar findings and suggestions regarding the need for training are noted extensively in the literature (Britton & Anderson, 2009; Kohut, et al., 2007; Wilkins, et al., 2009; Zundert, Sluijsmans, & Merrienboer, 2010).

A danger of peer review exists in what Marshall (2005) refers to as a “culture of nice.” Peers can sometimes be reluctant to provide constructive feedback and instead focus
their energy with pats on the back and simple praise. Building a culture of trust and respect can combat a culture of nice. Without trust, the positive outcomes associated with peer review may not be realized (Britton & Anderson, 2009; Byrne, et al., 2010; Iwanicki, 2001; Kohut, et al., 2007). In a survey study of peer review in higher education Kohut, Burnap, and Yon (2007) even advocate for establishing professional ethical guidelines to assist with the development of trust. In the survey of 223 respondents, they found that neither observers nor observees experienced significant stress during review, and many also reported that they learned as much from observing and providing feedback as being observed and receiving feedback. Iwanicki (2001) also saw reports of the same phenomenon in a study of 64 pre-service elementary teachers at 20 elementary schools. The participants were trained in giving feedback to peers, then asked students to independently develop a system for peer review with a classmate. Pre and post survey results were compared and the researchers found that after engaging in peer review, participants frequently reported that they became better teachers through their roles as observers as well. Participants also commented on the placing a higher value on collaboration, becoming more confident, developing a better understanding of children. However, the experience level of these pre-service teachers compared to the experienced staff at Mountainview Unified needs to be taken into consideration. The benefits of learning more about children and building confidence may have an amplified effect with pre-service teachers engaging in peer review as compared to veteran teachers. However, Iwanicki (2001) noticed similar results in a 2 year longitudinal study of rural elementary teachers. In this study, the peer review process increased teacher ownership of student
learning as well as valuing the quality of reflective judgment developed through collaboration.

Another potential benefit of peer review noted in research that may not be immediately obvious is the normalizing effects surrounding accountability when peers are exposed to each other's work and thinking. Fenwick (2005) notes that writing and analyzing goals in teams can coerce teachers into aligning their own goals with school goals. Weems and Rogers (2010) claim that peers help to establish and enforce standards of excellence. In a case study of nursing school faculty and administrators in Canada, Toth and McKey (2010) found that peer review could increase the consistency of the curriculum. This could be a substantial benefit for Mountainview Unified, which seeks to align evaluation with consistency of implementing initiatives outlined in our strategic plan. However, one limitation to the generalizability of these results is that the faculty reported the benefit of peer review acting as a buffer to possibly negative student course surveys. This added incentive might have affected the commitment to implementing peer review that would not be present in an elementary setting.

One arena where peer review has been used in elementary education is with peer assistance and review (PAR). In the early 1980's, school districts in Ohio worked closely with teachers' unions to develop systems of evaluation where mentor teachers worked with novice and struggling teachers to improve their practice. PAR was found to decrease attrition among new teachers because they felt supported and connected to their colleagues (Weems & Rogers, 2010). However this practice has not been widely used across the country and is often only incited when a teacher is struggling. In California,
teachers on improvement plans are often offered support through peer assistance and review as a last resort.

Marshall (2005) writes “The engine that drives high student achievement is teacher teams working collaboratively,” (p. 730) supporting the claim by citing such renowned educators as Richard Dufour, Mike Schmoker, Robert Marzano, Doug Reeves, Jeffrey Howard, Grant Wiggins and Jay McTighe. Although much of the research on peer review does not take place in elementary schools, the commonsensical findings of these closely related studies are generalizable to the elementary setting. It makes sense that anxiety would be lowered when being observed by a peer rather than a boss, allowing for more open dialogue and willingness to take risk. This would naturally allow for a sharper focus on collaboration and improving instruction than the traditional model of administrator drop-ins. However, care needs to be taken to establish a culture of trust and respect while also taking time to train staff on how to effectively participate in peer review. Such an environment can lead to increased satisfaction with the process with learning occurring for both observers and observees. However, the system of peer review needs to be thoughtfully developed to maximize these benefits.

Feedback

Having discussed the elements of a portfolio and processes to support evaluation, attention must be paid to how evaluative comments are delivered to teachers. Frequent and timely observation and feedback is critical to systems of evaluation (Britton & Anderson, 2009; Capizzi, et al., 2010; Conley, Muncey, & You, 2006; Fenwick, 2001; Marshall, 2005, 2008; The New Teacher Project, 2010; Weems & Rogers, 2010; Wilkins, et al., 2009). Most evaluation systems focus on year-end judgments representing a very small
fraction of the actual teaching and learning that occurs in the classroom. In studies of peer review, many of the researchers previously noted emphasized the importance of multiple observations and a cycle of interaction allowing for ongoing feedback (Byrne, et al., 2010; Kohut, et al., 2007; Wilkins, et al., 2009; Zundert, et al., 2010). However, recommendations range from providing feedback annually to once every one to two weeks. Observing a teacher for one 45-minute period of an academic year amounts to 0.1% of his or her actual teaching (Marshall, 2005). Weems (2010) states that even observing for three hours is not enough to build an accurate representation of a teacher’s abilities. Marshall recommends 12-15 visits per year, many unannounced, to help build a cycle on continuous analysis and reflection on one’s practice. By arriving unannounced, administrators and peers can gain an authentic perspective on day-to-day teaching rather than a teacher’s absolute best show that is thoroughly prepared for a planned observation. Sawchuck (2009c) advocates for three annual observations by an administrator and two observations by peers or master teachers. Multiple reviewers can help ensure consistency and accuracy of evaluations and avoid bias.

Once a decision has been made on the frequency of observations and providing feedback, the content and form of the feedback must also considered. Currently, some teachers simply receive “unsatisfactory” or “needs improvement” ratings as feedback without elaboration (Feeney, 2007). However, feedback must promote reflection and learning (Barbour, 2003; Feeney, 2007; Marshall, 2005; Siens & Ebmeier, 1996). A survey conducted at with faculty at an eastern university found that tenured and non-tenured faculty prefer narrative feedback as opposed to checklists or scores (Kohut, et al., 2007). In the case study of 15 teachers in large western school districts, researchers found that
suggestions and recommendations alone do not always engage teachers in reflective inquiry and lead to change (Feeney, 2007). The study defined quality feedback as using descriptive, observable data, providing characteristics of effective teaching, and promoting reflective self-inquiry supported by evidence of student learning. Based on the results of the study, Feeney recommends providing focused feedback, using structures such as questioning to promote reflective inquiry, and continuing to experiment with the evaluation process over time. An action research cycle directly supports the recommendation to experiment with the evaluation cycle on an ongoing basis.

Feedback can serve as both the disconfirming data and the new knowledge necessary to promote change and learning in teachers. However, barriers to effective use of feedback must be considered. Barbour (2003) warns that without appropriate trust, observers and observees can participate in collusion and be unwilling to take risks by providing frank and honest feedback. They remain safe and ignorant. However, there is also a danger of destructive openness where teachers’ vulnerabilities are not recognized, which can also shut down learning. Trust must be established so that teachers can let their guard down and feel psychologically safe so that they can access new learning.

**Clearly Defining “Good Teaching”**

Until this point, much discussion has occurred regarding structures and process of evaluation, but questions arise as to the processes used to benchmark against as well as the criteria used. The research and suggestions are mixed. The most widely accepted model of strong teaching is Danielson’s Framework for Teaching. Many school systems and teacher education programs throughout the country have adopted this model. Danielson claims, “To assess the quality of teaching practice, it’s essential to define it,” (2010, p. 2). However,
there is danger in adopting any particular framework of effective teaching. These concepts influence all aspects of teaching and can give teachers permission to ignore contradictory new information or practices that do not align with their belief systems (Young & Kim, 2010). Teachers have their own metrics of evaluating their effectiveness that do not align with the metrics of external accountability systems. Darling-Hammond, Pease, and Wise (1983) also note that there is little evidence of discrete, measurable, and specific actions that lead to increased student achievement. Instead, broader patterns of more nebulous traits such as clarity, enthusiasm, or task-orientation have found to have some effect, but even these have variable differences in diverse grade levels, subject areas, and settings. This speaks to the need to fully understand the specific organizational setting as well as beliefs on teaching and learning. To further confuse the issue, the same teaching strategy may have positive or negative effects depending on the goals of instruction. Particular strategies aimed at promoting achievement on standardized tests may have adverse effects on problem-solving and creativity. This research suggests that differentiating evaluation systems to meet the needs of individual organizations is critical. There is no silver bullet, no one-size-fits-all program that can quickly be adopted and administered. There are no context-proof, teacher-proof, or student proof actions and strategies. The complexities of institutions must be understood, illuminated and built upon.

Care needs to be taken to ensure that evaluation instruments and processes do not get in the way of teaching and learning (Marshall, 2005). Marshall seeks to simplify evaluation criteria using the acronym SOTEL. Observers should look for evidence of safety, objectives, teaching, engagement, and learning. Fenwick (2001) argues that reducing knowledge to observable, prespecifiable and trainable items subjects them to a sense of
perpetual deficit. Hargreaves (1998) expands on this idea of the dangers of a knowledge society that focuses on finding the one right answer or the one good way to teach. He says, “Until teaching is perceived, inter-alia, as a profession in which creative and adventurous but hard-headed pioneers feel at home, the negative image of the profession will persist,” (11). Interestingly, Hargreaves uses the word “profession” in this quote. Darling-Hammond, Wise, and Pease’s (1983) conception of teaching as a profession advocates for acknowledging the use of judgment and theoretical knowledge. Hargreaves is also advocating for creativity and adventure, core components of teaching as art.

However, Darling-Hammond (2010) also acknowledges that numerous studies have found that teachers’ participation in standards-based performance assessments can help teachers to improve their practice. A study of 178 high school teachers in southern California found that if teachers agree with the professional standards they tend to be satisfied with the evaluation process (Conley, et al., 2006). The more teachers understand the standards and see them as relevant, the more satisfied they are. Further satisfaction is also seen when teachers view the evaluators as trained and knowledgeable. A critical implication here is that increasing teacher participation in the development of standards and criteria on which to be evaluated can develop positive perceptions of evaluation. Involving teachers through action research embraces the concept of teaching as a profession and teaching as art.

**Conclusion**

Teacher evaluation is key in meeting the needs of future students.

Meeting the expectation that all students will learn to high standards will require a transformation in the ways in which our education system attracts, prepares, supports, and develops expert teachers who can teach in powerful ways – a
transformation that depends in part on the ways in which these abilities are understood and assessed. (Darling-Hammond, 2010, p. 2)

Using action research to explore the practices set forth in this literature review will provide teachers with the opportunity to co-create a system of teacher evaluation that allows them to learn and grow. Evaluating best practices through the lens of a specific conception of teaching while incorporating change theory will allow MUS teachers to develop a system of evaluation that meets their particular needs.
CHAPTER 3

Introduction

The purpose of my research was to develop a system of supervision and evaluation that meets the intent and statutory requirements of the Stull Act while incorporating an evaluation process that supports learning. A group of eight Teachers engaged in action research and their satisfaction with the co-developed teacher evaluation process was measured. Recent press on value-added assessment of teachers and the impact that teachers have on student achievement have placed teacher evaluation in the spotlight of education reformers. For many years, educational journals have unanimously endorsed the sentiment that national practices of teacher evaluation are generally ineffective. By involving teachers in action research to define their conceptions of teaching and co-create a system of evaluation, I hoped to increase satisfaction and the perceived utility of the evaluation system.

Research Questions

My research sought to address the following research questions:

1. What are the steps in the process of teachers co-creating a teacher evaluation system?

2. What do teachers at Mountainview Unified School consider to be important in evaluating teachers?

3. What do teachers at Mountainview Unified School believe are appropriate ways to evaluate teachers?
4. After participating in an action research model to develop a teacher evaluation system, how satisfied are teachers with the co-developed teacher evaluation process?

5. What are the challenges and lessons learned in the process of teachers co-creating a teacher evaluation system?

**Research Design**

This study employed a qualitative action research design in order to work with teachers, co-constructing a research-based system of teacher evaluation. The use of action research allowed me to document the teachers’ beliefs about teacher evaluation (what should be measured; what is important; and what are appropriate methods) as well as their perceptions of the process, lending itself to constructivist or participatory knowledge claims.

I studied the collaboration and creation process and perspectives of those involved. Action research allowed us to utilize an iterative cycle of inquiry, action, reflection, and revision. We began by conducting initial interviews to collect data on teachers conceptions of teaching as it relates to themselves and the school as well as their experience with, and thoughts about, teacher evaluation to establish baseline data. The group then reviewed literature on best practices in teacher evaluation, developed and implemented a model, reflected on the implementation, and revised it. Research questions were addressed by documenting perceptions and feelings throughout the process using field notes and interviews. The action research model aligned with my assumption that teacher evaluation systems should reflect individual institution’s beliefs about teaching as well as organizational culture. Action research uses non-researcher participants that have inside
knowledge of an organization as well as practical knowledge and expertise about the environment in which the study takes place. Furthermore, a qualitative approach utilizing observation and interviews allowed me to explore and investigate what MUS teachers feel is important in teacher evaluation; what methods of evaluation are appropriate to our unique organizational context and population; satisfaction with the process; and ultimately, steps, challenges, and lessons learned so that others may benefit from this research. In particular, personal interviews allowed me to probe individual's beliefs and perceptions throughout the process of creating a system of evaluation. Field notes and transcripts from group discussions allowed me to better understand what the culture of MUS values in regards to teacher evaluation.

As process and perspective are the focus phenomena in my study, a quantitative design that employed surveys or experiments would be inappropriate. A purely quantitative study would not afford me the opportunity to fully explore teacher beliefs, feelings and perceptions. Because we sought to co-create our understanding and ultimately a new system of evaluation, it was important to document the detail and process of construction to better understand and be able to describe why certain outcomes occurred. Quantitative research generally describes outcomes, but is not flexible enough to probe for explanations. The small sample (n=8) also made any quantitative analysis beyond reporting frequencies inappropriate.

A purely qualitative study without action research would not have enabled me to answer the research question of how participation in the process of creating a system of teacher evaluation can lead to satisfaction. An underlying theory in my literature review and research design is that for learning to occur, efforts must be made to establish a culture
of trust and risk-taking while also involving teachers in the process. A qualitative study without action research would not have met these needs and I would have been unable to answer the research questions.

**Site Selection**

The research took place at Mountainview Unified School (MUS), a public K-6 elementary school in Montecito, California serving 440 students. This site is unique in that Montecito is one of the most affluent areas not only in California, but the nation. Because California schools are funded through local property taxes, MUS is also one of the most well funded schools in California, with a budget of nearly ten million dollars annually. MUS spends nearly $23,000 per student compared to other districts that might spend only $5,000 per student. The students and teachers are primarily white, with other ethnicities comprising an additional ten percent of the population. The teaching staff is also unusually veteran, with an average of 20 years of experience.

Another unique structure of this school is the employment of specialists such as art, music, physical education, technology and Spanish that allows teachers to leave their classes, freeing them up to observe one another, prepare materials, or collaborate in teams. This is a critical difference as it relates to a teacher evaluation system potentially involving peer evaluation. Also, Mountainview Unified has an excellent reputation and is designated as a “10/10” school, performing in the top decile of all California schools as well as the top decile of 100 “similar schools” based on demographic characteristics. However, in the Mountainview Unified School District, the evaluation process has fallen below the minimum benchmarks set by the Stull Act. At a staff meeting in September of 2010, a

---

4 The Stull Act requires school districts to evaluate and assess the performance of certificated
teacher who had taught at the school for over 15 years commented, “I don’t think I’ve ever been evaluated.” A review of personnel files indicated that between 2002 and 2010 no teacher had more than two evaluation documents on file. These documents generally contained three to four sentences with ambiguous statements such as “The teacher differentiated instruction for multiple levels.”

All of these characteristics made MUS a unique site for investigation. Furthermore, the availability of many resources allowed for development of a near ideal system with few restricting factors.

As the Chief Academic Officer at Mountainview Unified, I was in a unique position to gain access to the site. My primary role is to oversee teaching and learning, including the supervision of all credentialed staff. With the help of the superintendent, I have worked hard to create an atmosphere of trust and respect through the process of developing team norms, prioritizing the school’s strategic plan, improving communication systems, and embracing a code of honesty, transparency, and collaboration in all matters. This helped me be perceived as a person who can be trusted and gave me emotional access to the participants.

Sample

My study included eight teaching staff members at Mountainview Unified School. The school employs 31 full-time credentialed teachers. Of these, 94% are white and four are male. Staff members include kindergarten through sixth grade teachers as well as specialists in art, music, physical education, math, reading, Spanish, and science. Teachers have experience ranging from five to more than thirty years. The staff is also unique in that employees that perform the requirements of educational programs mandated by state or federal law (Education Code Sections 44660 to 44665). See Appendix A for complete details.
their collective bargaining agreement contains language mandating that they be the highest compensated teachers in the state of California. Many teachers make over $100,000, with the top teacher earning $116,000.

I used opportunity sampling to select teachers from the population. The purpose of using action research is to utilize the expertise of the teachers and tailor the program to fit their needs. I opened participation to all teachers wanted to ensure I had at least one representative from each span of two grade levels (i.e. K-1, 2-3, 4-5, and 6) as well as a representative for the specialists. I invited all staff members at the site at a staff meeting as well as in writing through the staff bulletin. The expected time commitments were outlined and the consent to participate was thoroughly discussed to avoid coercion. It was made clear that participants were free to withdraw from the study at any time should they choose to participate.

After the first invitation was sent on October 10, 2011, I did not have a second or third grade teacher, so a specific invitation was sent on October 14, 2011 to the grade levels and one teacher volunteered. I planned to randomize selection by drawing names if too many applied in order to remove any bias, but only eight volunteered. Keeping the sample below ten allowed for more active participation in group settings yet also provided a representative sample of approximately one fourth of the population. Those who volunteered closely approximated school ratios in relation to experience, ethnicity and gender. Eight staff members volunteered and the consent document was thoroughly reviewed once again prior to obtaining signatures on October 19, 2011. A range of grade levels from kindergarten through 6th grade were represented in this group as well as a range of experience levels. Of the eight participants, six were female and all were white,
closely mirroring the overall demographics of the school, where 92% of the certificated staff members are female and white. Experience levels were diverse, ranging from a teacher within her first 5 years of teaching to two teachers retiring at the end of the academic year in which the study took place. Specific demographic characteristics are described in the table below.⁵

Table 3.1: Participant Demographic Data

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>Years Teaching</th>
<th>Years Teaching at MUS</th>
<th>Grade Level/Subject</th>
<th>Highest Degree Obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathy</td>
<td>Female</td>
<td>White</td>
<td>7</td>
<td>2</td>
<td>K</td>
<td>Masters</td>
</tr>
<tr>
<td>Tracy</td>
<td>Female</td>
<td>White</td>
<td>46</td>
<td>28</td>
<td>2</td>
<td>Masters</td>
</tr>
<tr>
<td>Loretta</td>
<td>Female</td>
<td>White</td>
<td>17</td>
<td>15</td>
<td>4</td>
<td>Masters</td>
</tr>
<tr>
<td>Isabelle</td>
<td>Female</td>
<td>White</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>Masters</td>
</tr>
<tr>
<td>Nathan</td>
<td>Male</td>
<td>White</td>
<td>23</td>
<td>23</td>
<td>4</td>
<td>J.D.</td>
</tr>
<tr>
<td>Ethan</td>
<td>Male</td>
<td>White</td>
<td>25</td>
<td>18</td>
<td>5</td>
<td>Masters</td>
</tr>
<tr>
<td>Nancy</td>
<td>Female</td>
<td>White</td>
<td>43</td>
<td>26</td>
<td>6</td>
<td>Bachelors</td>
</tr>
<tr>
<td>Sarah</td>
<td>Female</td>
<td>White</td>
<td>6</td>
<td>3</td>
<td>Music</td>
<td>Bachelors</td>
</tr>
</tbody>
</table>

Because Schein (1995) states that psychological safety must be created in order for people to be able to take risks and ultimately learn, the superintendent, board members, parents and students were specifically excluded from this study. Teachers will be more willing to take risks and learn if surrounded by peers, rather than worrying about the perceived impressions of ‘outsiders’. Furthermore, they have the necessary skill set and knowledge to develop a system of evaluation that is meaningful to them. A vocal parent or student might strongly advocate for a specific aspect of evaluation that fits their personal need or bias, but is not necessarily grounded in best practice. Because one goal of this

---

⁵ Table 3.1 is sorted by grade level and pseudonyms are used to protect the confidentiality of the participants
process was to create a system that teachers feel ownership of and focuses on their learning, the study involved teachers only.

**Access**

The superintendent of MUS agreed to provide generous access to the site and its resources for the purpose of this study. In addition to general permission to conduct the study and implement the pilot, I was granted full access to any district resources that do not incur monetary expenses within reason. I had access to the school’s email system, phones, mailboxes, bulletins, and newsletters for communicating. Staff meeting time was used to introduce the project, recruit participants, and deliver logistical information throughout the research. Teachers were accessible during contract time such as lunch, breaks, or when their students were at specialists (art, music, physical education, etc.). I also had permission to use my time during the work day to conduct interviews, observe teachers, and coordinate the research. However, the school did not support any extra compensation for non-contract time or purchasing of materials not already on-site. I also was not permitted to write my dissertation on contract time.

**Data Collection Methods**

I sought to answer the research questions with the following methods:

**Table 3.2 – Data Collection Methods**

<table>
<thead>
<tr>
<th>What do teachers at Mountainview Unified School consider to be important in evaluating teachers?</th>
<th>Field observation notes, journal responses, interviews</th>
<th>Frequencies of items reported:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Purpose of evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Student achievement data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Motivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Standardized measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Portfolios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Feedback on student work</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| What do teachers at MUS believe are appropriate ways to evaluate teachers? | Field observation notes, journal responses, interviews | • Peer observation  
• Assessment results  
• Video  
• Rubrics  
• Frequency of observations  
• Professional teaching standards |
|---|---|---|
| After participating in an action research model to develop a teacher evaluation system, how satisfied are teachers with the co-developed teacher evaluation process? | Field observation notes, journal responses, interviews | Statements of overall satisfaction with process  
Reports of:  
• increased collaboration  
• Diminished isolation  
• Valuable use of time  
• Appropriately challenged  
• personal growth and improved practice  
• feeling supported  
• low anxiety  
• Increased test scores  
• Peer expertise  
• Trust  
• Inspired to improve  
• Sense of ownership  
• Sense of involvement  
• Satisfaction with the created instrument/process |
| What are the steps, challenges, and lessons learned in the process of teachers co-creating a teacher evaluation system? | Field observation notes, journal responses, interviews | • Plan, do, check, act  
• Time together  
• Supervisor influence  
• Authentic conversation  
• Professional learning communities  
• Defining conception of teaching  
• Using research |
**Rationale for Methods Used Table 1:**

<table>
<thead>
<tr>
<th>Journal Response – Teachers anonymously wrote response journals on the action research process, ultimately reflecting on what they valued in the developed evaluation model. Anonymity was ensured by removing all identifying marks from the entries, but marking by pseudonym to connect writings to particular participants throughout the research. Participants were allowed to handwrite or type entries. Journal entries were in response to specific prompts. Personal journals provided insight into individual teachers thoughts outside of group interactions.</th>
<th>Interviews – Individual interviews of participants allowed me to ask common questions to all participants while also allowing me to probe for thorough responses on why they were or were not satisfied with the different responses to the survey. These interviews were conducted in person on site. One-to-one interviews reduced participant reactivity.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Field observation notes</strong> Throughout workshops and discussions, teachers had the opportunity to co-construct new learnings through their interactions. It was important to document this process in order to get a strong understanding of group beliefs and the climate of interactions among teachers. This provided documentation of what they believe is important in teacher evaluation and how they are evaluated.</td>
<td></td>
</tr>
</tbody>
</table>
The timeline below describes the sequence of events employed throughout the study.

**Figure 3.1: Timeline of Events**

Data collection began with the entire staff of 36 credentialed teachers being given the opportunity to complete a pre-survey sent out on October 19, 2011, (Appendix B) asking questions about teachers experiences and beliefs about the purposes of teacher evaluation and best methods used to evaluate teachers. Only four of the staff completed the pre-survey, and all four were members of the group of eight participants. Their responses can be viewed in Appendix J. To avoid coercion, I did not place pressure on the staff for additional responses, and thus have not included the pre-survey data as part of my findings.

Data collection continued with 30-45 minute personal interviews for each participant. These took place from October 24, 2011 to November 2, 2011. The interview protocol (Appendix C) consisted of open-ended questions regarding beliefs and perceptions about teaching, teacher evaluation and how these relate to the specific context of Mountainview Unified School. Specifically, Darling Hammond, Wise and Pease's (1983) continuum of teaching was presented and participants were asked to approximate their
beliefs as well as their peers as they related to the continuum. These initial interviews helped me collect background and baseline information as well as explore beliefs around teaching and evaluation. All meetings and interviews were audio recorded with explicit permission from all participants. Transcriptions were made available to them for review in the event that they wanted to clarify statements or request that material be omitted. Interviews took place in a private administrative office. Meetings took place at the school site in the teacher’s professional learning community classroom, a room dedicated for teachers’ use in reflecting on and improving their practice.

After the initial interviews, participants read a review of literature on best practices in teacher evaluation (see Chapter 2). They were given prompts (Appendix D) and asked to record their thoughts. Prompts asked them to share new understandings of teacher evaluation after reviewing research, connect their beliefs about the continuum of teaching to evaluation practices, and consider which practices might align best in the specific context of MUS.

Shortly after the reading and journaling were completed, the first group meeting with all participants was held on December 15, 2011. At this 90-minute meeting, I shared patterns and trends from the initial interviews and journal responses. The group reflected on their beliefs about teaching and evaluation and worked to develop a rough outline of a pilot evaluation model to be implemented for six weeks. Field observation notes were taken and conversations were recorded and transcribed.

After the first group meeting, I worked to encapsulate what was said and map out the details of how what was expressed in our discussion could be turned into a working
system. After reviewing my notes and listening to the audio recording of the meeting, I created the plan outlined in Appendix H and presented it to the group.

To create the plan, I used the findings above and agreements made throughout the meeting. We had clearly agreed to a menu of evidence that would be gathered in a portfolio. From this information, I outlined the details of the plan and included two sample plans that I created (Appendix H) to provide a lens for the scope of work I was envisioning, setting a standard for quality and rigor. I also provided sample questions to guide the content of reflections as well as the mid-cycle study group meetings.

Next, we reviewed what I had come up with together at a second group meeting on January 4, 2012 and agreed to implement the plan with one small change. Rather than pre-conference with me alone, the group thought it would be helpful to pre-conference with their study group to get feedback from peers as well as an administrator, reflecting their desire for peer interaction.

Once the evaluation system was agreed upon, the group of eight split up into two smaller groups randomly. Group A consisted of Isabelle, Loretta, Ethan and Tracy while Group B included Nathan, Nancy, Cathy, and Sarah. These two groups then met with me separately (January 5 and January 9, 2012) and we reviewed their ideas together. Questions were asked, clarifications and suggestions were made, and professional growth plans were solidified.

We then began implementing the pilot according to teachers’ professional growth plan. For some, this consisted of collaborating with a specific group of teachers while others worked with a partner or independently. Many teachers invited me to in to observe lessons while I stopped by frequently and unannounced for others. Some stopped by my
office to think together about next steps or bounced ideas off of colleagues. After four weeks of moving forward in this direction, the two study groups met individually during the week of February 6, 2012 for a check-in, accompanied by a short reflection answering the question asking them to consider how their work so far had impacted student achievement. At these 30-minute meetings, teachers reflected on what they had tried so far and offered advice and encouragement to one another. They also spoke about the process itself and its relative pros and cons to a traditional model of pre-announced observations. The groups were asked how much more time they needed to feel comfortable with their progress towards their goals as well as the evidence collection process. Both groups agreed that an additional two weeks would be sufficient.

After the two weeks passed, the study groups met with me as two separate groups to share their evidence and turn in their final reflections. The meetings lasted approximately an hour, giving each participant 15 minutes to discuss their evidence and growth as a teacher as well as impacts to student achievement. Some teachers put together 3-ring binders while others had loose collections of work and digital media such as photos and videos.

Approximately one week after the portfolio sharing meetings, seven of the eight participants met with me for an hour to reflect on the process as a whole (Nancy was ill). The group was asked to make recommendations regarding the continuation of the project at MUS as well as consider factors that might influence adoption of the evaluation system or process at the local, state or national level.

Finally, the study was completed by conducting individual interviews with each participant in the days immediately following the final group reflection during the week of
March 5, 2012. Open-ended questions regarding beliefs and perceptions about teaching and the teaching continuum, teacher evaluation and how these relate to the specific context of Mountainview Unified School were used. Individuals answered question about their satisfaction with the process, why they were or were not satisfied, and how, if at all, their beliefs had changed over the scope of the project. Interviews were conducted in my office and lasted approximately thirty minutes.

After examining the findings in depth, it was noted that there was a lack of specific information regarding the use of standardized testing from participants. Additional individual interviews were conducting during the week of April 2, 2012 to explore participants beliefs on standardized testing and the use of test results in teacher evaluation.

All interviews and group meetings were recorded, transcribed, and made available for participants to review. Audio recordings were transferred from the digital recorder to a laptop and backed up on an external hard drive. The interview and group meeting transcripts, journal responses, and field observation notes along with their analysis were stored in both electronic and paper form. All electronic documents were password protected and hard copies were kept in a binder under lock and key.

Data Analysis

Data collection and data analysis are often considered simultaneous tasks. It is difficult to do one without the other (Merriam, 2009). The three methods (interviews, field notes, and journal responses) helped triangulate the findings. As a “participant as observer,” the field observation notes were subordinate to my role as participant-facilitator. Journal responses and interviews were transcribed and analyzed for patterns and themes in order to develop categories for coding and dumping data. Main categories
included beliefs about teaching, beliefs about evaluation, and beliefs about the action research process. These main categories were then sectioned into more specific sub-categories that arose from the data (See Appendix K). For example, when exploring beliefs about evaluation, sub-categories included videotaping, portfolios, peer observation, student work and more. These categories developed naturally by analyzing interview transcripts, field observation notes, and journal responses for patterns and trends.

Once the coded responses were sorted, I developed frequencies of responses as well as anecdotal evidence to explain their thinking. Responses were linked to their definitions of teaching through Darling-Hammond’s framework. The coded findings were made available to participants to ensure their statements and intentions were accurately represented. This approach of formulating theories that arose from data-based themes is designed to strengthen the validity of the study (Maxwell, 2005).

Findings from interviews, journals and field notes were utilized throughout the action research cycle to inform interview and group questions as well as journal prompts. While analyzing data, I journaled myself, making note of themes and areas to explore further. Specifically, I analyzed how participants’ beliefs and perceptions changed over time throughout the study, how beliefs, feelings and decisions may have been explained by theory, and overall satisfaction with the process and outcome.

**Ethical Considerations**

Significant ethical considerations existed in this research but were mitigated by having proper systems in place. Action researchers need to be aware of bias, as it can be difficult to step back and look at one’s own institution objectively. Action researchers also need to be particularly careful with role duality, balancing duties as practitioners and
researchers. Despite having improved access in many areas, inside researchers such as myself may have limited access to objective information. Participants were aware of political, cultural, and social implications of their actions and responses. As the Chief Academic officer and their supervisor, there was the potential for teachers to feel coerced into participating in the study or into providing answers, actions, or system designs that they thought I wanted based on our previous interactions. Teachers were asked to sign clear consent forms that acknowledged the voluntary nature of the study. These forms also clearly expressed that no one would be disciplined or penalized for removing him or herself from the study, working to eliminate any coercion. Participants were given another reminder during initial interviews. When managing my role, I made efforts to regularly remind teachers of my role as a researcher during meetings and interviews, and downplayed my role as their supervisor.

Care was given to ensure confidentiality throughout data collection. Non-identifiable markers were used in place of names in all data. Audiotapes of meetings were used instead of videotapes to provide one more measure of anonymity. Data was kept in a locked file at all times and electronic data was password protected.

I also needed to be careful not to allow my own biases and background knowledge of the site and teachers to influence the research. I endeavored to approach each task with impartiality, asking myself for evidence of my conclusions and journaling my reactions throughout the process so that I could later reflect on the neutrality of my thinking.

Participants in this study may have been exposed to risk in that peer observation was an outcome of the action research and evaluation design. Teachers were at risk of being perceived negatively by peers and having feelings of anxiety caused by observation.
These risks were reduced by developing a clear understanding of confidentiality protocols and establishing a climate of risk-taking through shared reading of the literature review.

Another important consideration in the ethics of this study is the fact that a well-designed system will benefit the teachers over the long term. They had a vested interest in maintaining the quality of the study because the outcome may determine the system of evaluation that is utilized at the site in the future. Furthermore, this cycle of research is commonly used for many decisions at the school level where teacher input is sought. The administration embraces collaborative decision-making based on best practices supported in research and uses a similar structure on a regular basis. Because the format of this particular study was so similar to typical work of the school, participants could be expected behave in typical fashion, thus supporting authenticity of the data.

**Reliability and Triangulation**

To ensure accuracy, I conducted member checks (respondent validation), by asking interviewees to review conclusions drawn from transcript analysis and data coding. This ensured that the meanings of interviewee statements were interpreted correctly while helping me identify my own preconceptions and misunderstandings.

Accuracy and reliability were enhanced by using journal prompts and interview protocols that were reviewed and piloted. The review process entailed a thorough review by my dissertation chair and peers involved in similar work. Instruments were also piloted and refined with non-participants. Piloting the instruments helped ensure that questions would not be misinterpreted or unclear. (Fowler, 2009). Results of the pilots were used to clarify language and design elements. On key themes, the journal prompts and interviews also asked the same question in multiple formats to rule out any potential misreading of
individual questions. Reactivity was also mitigated by frequent contact with participants. They were less likely to say what they believed I wanted to hear through frequent interactions and reinforcement of norms. However, most important to reducing reactivity was reinforcing the belief that I wanted their authentic input and was committed to building a system that works for them. We used the school’s current norms, which required us to make every decision in the best interest of students. Their prior interactions with me, in which I have strived to always put the needs of the students and staff above my own personal needs, helped to build my credibility, thus limiting reactivity.
CHAPTER 4

Introduction

This study aimed to document and analyze the process of co-creating a system of teacher evaluation with public elementary school teachers, to identify what criteria teachers should be evaluated on, and to determine appropriate methods of evaluation. An action research approach was used to develop and implement a system of evaluation focused on teacher growth. A group of eight elementary teachers met from November of 2011 until March of 2012 to share their beliefs about teacher evaluation, study research in best practices, and build a model of evaluation that aligns with their beliefs and resources. Data were collected through interviews, journal responses, and group meetings. The study took place at a highly affluent and well-funded public elementary school with the highest state test scores in its southern California county.

This chapter will describe the process used to develop and pilot the system while identifying common themes and findings that emerged throughout the process, seeking to answer the following questions:

1. What are the steps in the process of teachers co-creating a teacher evaluation system?

2. What do teachers at Mountainview Unified School consider to be important in evaluating teachers?

3. What do teachers at Mountainview Unified School believe are appropriate ways to evaluate teachers?
4. After participating in an action research model to develop a teacher evaluation system, how satisfied are teachers with the co-developed teacher evaluation process?

5. What are the challenges and lessons learned in the process of teachers co-creating a teacher evaluation system?

In this chapter, I will explore the findings through the lens of the research questions. Because the cycle of action research was an integral part of this study, I will begin by discussing the first part of research question one, the steps, and their importance, in co-designing a system of evaluation with teachers. After developing the context, I will then discuss questions two and three, exploring teachers’ beliefs about evaluation in general as well as their thoughts on the specific mechanics of an evaluation system. Though teachers’ beliefs were more centered on teacher growth than student achievement, I will take the opportunity to address how this study impacted student achievement through these two questions. Finally, I will discuss teachers’ satisfaction with the system and return to the second piece of research question one, the challenges and lessons learned along the way.

The timeline below describes the sequence of events that occurred throughout the study.

**Figure 4.1: Timeline of Events**
Summary of Findings

The findings in this chapter are organized around the process of developing a system of evaluation, beliefs about teacher evaluation, methods of evaluation, and satisfaction. Throughout these findings, some common themes emerged. Participants clearly felt satisfied with the co-developed system of evaluation and believed it led to their growth as teachers, thus impacting student achievement.

The sub-findings of this research study can best be understood under the umbrella of the following major findings:

- Teachers believe that the purpose of teacher evaluation is to help them grow as educators. This can be accomplished by providing a safe and supportive culture and systems that motivate and engage while remaining flexible enough to meet the needs of diverse teachers.

- Teacher growth is supported through various methods, including peer interaction, goal setting, evidence collection, reflection and an ongoing and iterative process.

- Teachers reported being highly satisfied with the co-developed model of teacher evaluation and believed it led to teacher growth and gains in student achievement.

- Many lessons were learned along the way that can help others successfully replicate a similar process in their own schools and districts. Most importantly, specific logistical parameters were recommended, the importance of precise language and clarity was seen, and the process of reading the literature review changed teachers' beliefs about evaluation.
These major findings are presented in depth as they relate to the four research questions guiding this study.

**Findings**

**RQ1: What are the Steps in the Process of Teachers Co-Creating a Teacher Evaluation System?**

As I discuss the findings out of context, it is important to understand the data within the sequence of events that occurred throughout the action research cycle. To build a better understanding of the process, each of the seven stages are described below as well as the intentional nature of their order.

**Stage 1: Whole staff pre-survey and initial interviews.**

The entire staff of 36 credentialed teachers was given the opportunity to complete a pre-survey, asking questions about teachers experiences and beliefs about the purposes of teacher evaluation and best methods used to evaluate teachers. Only four of the staff completed the pre-survey, and all four were members of the group of eight participants. Their responses can be viewed in Appendix J. To avoid coercion, I did not place pressure on the staff for additional responses, and thus have not included the pre-survey data as part of my findings. The return rate may have suffered because of unfortunate timing and overlap with report card writing and parent conferences.

Open-ended individual interviews were conducted using the questions in Appendix B. Interviews took place in a quiet office on the school's campus before and after school or while a teacher’s students were at a specialist such as art or music. The interview protocol was designed to gather baseline information about the teachers’ beliefs about the purposes and methods of teacher evaluation as well as their past experiences.
As my literature review indicated, I operated on the assumption that for a teacher evaluation system to be effective, it must be tailored to what the local teachers believe about teaching and the purpose of evaluations. Darling-Hammond, Wise and Pease (1983) argue that a thorough understanding of context is critical in designing successful teacher evaluation. Common views of the teaching-learning process must be articulated in order to determine what the organization values and believes, which in turn helps determine how teachers are measured. By questioning teachers about their past experiences, how they would improve them, and what they believe about teaching and learning, common trends emerged that helped guide the development of a system around their beliefs.

**Stage 2: Literature review and journal responses.**

After the initial interviews were conducted and baseline data were collected, participants were asked to read the literature review (Chapter 2) and respond to journal prompts asking about new understandings and beliefs about evaluation. Participants were given two and a half weeks to read the literature review and respond to the prompts in Appendix D. Six out of eight teachers returned written responses. Though not all participants responded, patterns and commonalities arose. This step was important in exposing the participants to research on best practices in evaluation. Teachers were asked to incorporate their personal beliefs with knowledge of best practices to ensure a thoughtful and well-designed system of evaluation.

**Stage 3: Group discussion and system design and review.**

After interviewing teachers and having them read the literature review, we met as a group to discuss our beliefs about teacher evaluation and began to design a system. Staff members vocalized their beliefs and their thoughts and worked together to articulate key
attributes that a system at MUS should contain. Responses were recorded on chart paper and revised as necessary throughout the conversation. Thoughts and beliefs that were shared during interviews and journaling were reiterated and patterns emerged that lead to a design that they had all participated in developing. Implications of this process will be seen in the section on RQ4 where teacher satisfaction is discussed.

After the first group meeting, I worked to encapsulate what was said and map out the details of how what was expressed in our discussion could be turned into a working system. After reviewing my notes and listening to the audio recording of the meeting, I created the plan outlined in Appendix H and presented it to the group.

To create the plan, I used the findings above and agreements made throughout the meeting. The idea of a professional growth plan was based on Fenwick’s (2001) research that found that the use of professional growth plans developed a greater sense of commitment and focus among teachers. They felt that they were held more accountable, but also felt more collegiality.

We had clearly agreed to a menu of evidence that would be gathered in a portfolio. For the sake of this research project, I decided to base the portfolio on one goal rather than two as we discussed at our meeting. I did this because of the shortened eight-week window of the pilot implementation. As noted above, the teachers were very concerned about the time and resource demands of an evaluation system and I did not want to overwhelm them. The discussion on goal-setting was connected to Brogan’s (1995) claims that setting goals can help teachers begin with the end in mind and can provide structure to the use of portfolios.
Although it was not explicitly agreed to at the planning meeting, the teachers in this study regularly referred to the power of reflection in learning and growth in their interviews and journaling. We also saw earlier in the literature that Feeney (2007) claims learning cannot happen without reflection. This led me to include reflection as one of the two must-do activities. Peer interaction was another important aspect of learning that teachers vocalized in interviews, journal prompts and the group meeting. I designed the system to include study groups where approximately four teachers would check-in with one another on their progress mid-way through the pilot. They would then present their portfolios to this same group at the end of the pilot. The use of study groups was included in an effort to include what Danielson (2010) describes as a tangential benefit when using peer collaboration. Sharing the process as well as portfolios with one another can increase peer learning through evaluation of each other’s thinking and evidence as well as an increased sense of collaboration. Participants could potentially learn not only from their own experiences, but through exposure to their peers' activities. Finally, because the importance of clarity had come up numerous times, I included two sample plans that I created (Appendix H) to provide a lens for the scope of work I was envisioning, setting a standard for quality and rigor. In these samples, I also provided sample questions to guide the content of reflections as well as the mid-cycle study group meetings.

Together, we reviewed what I had come up with and the group agreed to implement the plan with one small change. Rather than pre-conference with me alone, the group thought it would be helpful to pre-conference with their study group to get feedback from peers as well as an administrator, reflecting their desire for peer interaction. This idea was sparked by teachers noting how powerful the sample plans were that I created. They were
powerful because they gave them a sense of the scope of the work as well as ideas for evidence. They also valued hearing multiple perspectives on their own work and expected to be inspired by hearing what others were doing.

An understanding of the pilot model is essential to making sense of the remainder of this chapter. The agreed upon framework can be seen in Appendix H. To briefly explain it, teachers began by meeting in study groups with three other peers and myself. At these meetings, they shared ideas for potential goals and means of collecting evidence with one another, giving and receiving feedback. After this meeting, they formalized their plans by writing out specific goals, evidence to be collected, the importance and relevancy of the work to their current teaching, and other agreements. Each teacher was required to include observation and written reflection as part of their evidence, but were free to collect any other evidence they felt would demonstrate progress towards their goals. They were then sent on their own to pursue their goals and implement their plans.

**Stage 4: Implementation of pilot via study groups.**

Once the evaluation system described above was agreed upon, the group of eight split up into two smaller groups randomly. Group A consisted of Isabelle, Loretta, Ethan and Tracy while Group B included Nathan, Nancy, Cathy, and Sarah. These two groups then met with me separately and we reviewed their ideas together. Group A volunteered to conduct their pre-conference the following day. Because of availability of the group members and myself, Group B did not meet until four days later. Questions were asked, clarifications and suggestions were made, and plans were solidified (see Appendix I for final professional growth plans). These professional growth plans became the driving documents for teachers’ learning process and documentation of growth.
Stage 5: Mid-implementation study group check-in.

After four weeks of implementation, study group participants met with me to discuss their progress towards goals and share preliminary evidence that had been collected. Participants were also asked to respond to a simple journal prompt, “What impact have your efforts toward fulfilling your professional growth plan had on student achievement?” Ethan was unable to participate in this stage of the research because of a family illness that prevented him from attending the meeting. The check-in also provided an opportunity to clarify expectations before the work got too far along and remind participants what they had agreed to.

Stage 6: Portfolio sharing and individual reflections.

At the end of the pilot session, study groups came together to share their evidence and progress towards goals. As defined earlier, teachers created portfolios, collections of evidence mapped out in their professional growth plans, including student work samples, written reflections, pictures, videos, lesson plans, and more. Some participants chose to put together a standard 3-ring binder while others created PowerPoint presentations and digital slideshows. This process provided the data necessary to evaluate the impact of the evaluation system as well as teachers’ satisfaction.

Stage 7: Group reflection and final interviews.

After the portfolio sharing meetings, seven participants reconvened to discuss the process as a whole (Nancy was ill). They were asked to reflect on the process and offer suggestions for the future of teacher evaluation at MUS. They were also asked to make more general recommendations regarding teacher evaluation for others to consider if there were interested in adopting a similar system or engaging in a similar method of developing
a system. Following the group reflection, each participant was also interviewed
individually to further explore their thoughts and beliefs regarding teaching and teacher
evaluation, why this system did or did not work, and their satisfaction with the process.

**Stage 8: Follow up interview on standardized testing**

Though not originally planned, one more round of ten to twenty minute individual
interviews were conducted in early April, 2012 to explore participants feelings regarding
standardized testing and use of test scores in teacher evaluation. This information was
needed to address current national debate on the use of value-added scores in teacher
evaluations.

**RQ2: What do Teachers at Mountainview Unified School Consider to be Important in
Evaluating Teachers?**

*Teachers believe that the purpose of teacher evaluation is to help them grow as
educators. This can be accomplished by providing a safe and supportive culture
and systems that motivate and engage while remaining flexible enough to meet the
needs of diverse teachers.*

As Darling-Hammond, Wise and Pease (1983) assert, a system of evaluation must be
aligned with local beliefs in order to be effective. By first understanding teachers’ beliefs,
studying research-based best practices helped inform the development of a system that
supports and promotes their conception of teaching. Teachers’ beliefs were explored
through individual interviews, written reflections, and field observation notes and
transcripts from group meetings. It was clear that teachers did not believe in value-added
measures, yet they did value student gains on other measures that occurred throughout
this study. They felt the purpose of an evaluation system should be to support teachers’
professional growth through a motivating and engaging structure. They asserted that the developed system must create a safe and supportive culture and be flexible enough to be individualized for the diverse needs of teachers.

**Teachers do not believe that value-added measures of student achievement and teacher rankings or grades are useful evaluation tools.**

In the preliminary planning stages and final group reflection and interviews, each participant clearly expressed the detrimental effects of relying on value-added measures based on standardized testing that rank teachers. As described in more detail below, teachers placed strong emphasis on creating a supportive and safe environment in which teachers can learn and grow. Much like Schein’s (1995) sentiments regarding psychological safety, participants articulated that they must feel safe in order to take risks. A culture of collaboration and support was essential to improving their practice. They felt strongly that using value added measures would create unhealthy competition amongst teachers.

In the final group reflection and individual interviews, participants elaborated on why the use of value-added scores is counterproductive to teacher growth. Ethan explained, “I think you sort of shoot yourself in the foot as soon as you say, ‘here's your value added score.’ ... It's all over.” He described how teachers would live in fear and stop taking risks. Nathan described what he saw as a lack of personalization to education when teachers at MUS were overly focused on test scores in prior years. He felt, “If you start doing the tit for tat thing on what children are in your class, I don’t want to know his [any student’s] personality. I don’t want to know what his life situation is. I want to know what his test scores are, because they are most important to my success.” He went on to
describe how this culture forced teachers into “seeing your colleagues as competitors for a
valuable commodity, high test score students.” Isabelle also viewed value-added scores
negatively, adding, “It’s more counterproductive than it is productive ... I think it’s a
slippery slope ... because then everyone kind of gets that competitive edge and it turns into
a negative really quickly.” Each participant agreed that assigning value-added scores to
teachers or ranking them would inhibit risk-taking and a supportive environment where
teachers feel safe. This is critical to understand considering Schein’s (1995) premise that
adults must feel psychological safety in order to change. We sabotage the system by
publishing test scores and eliminating this safety. Ethan commented on evaluating
teachers via value added scores, saying, “You don’t need a stick. You know, we’re not even
giving you a carrot.”

In addition to creating an unhealthy climate of competition and discomfort,
participants felt that current standardized tests do not necessarily measure what is most
important in a child’s education. In particular, measuring teachers’ standardized test
scores does nothing to identify specific aspects of great teaching. Ethan voiced his
frustration with current state assessments in the final group reflection, saying, “We’ve got a
bad system. We don’t think it really evaluates kids the way that they need to do it. It
doesn’t evaluate good teaching, it doesn’t evaluate good education.” Tracy built on these
thoughts, sharing,

If your goal is to have children who think and problem solve and think about their
thinking and communicate, then you’re going to want an evaluation of teachers that
shows they’re promoting that. If you want kids that can compute at a certain speed,
who know how to color in the lines and march in a straight lines, then you have a
different evaluation system for that. And some of the competing things in the news
that you’re hearing, people want to rate the teachers based on their performance,
and their idea of what a good teacher is really skewed from mine.
She later added in her final interview, "I can’t get a number on how excited they were about the way they learned, the things they told their parents, the things that they, you know, produced in other ways.” It was clear from their comments that there were much more important aspects to their work than standardized test scores, and felt that the evidence demonstrated in their portfolios and stories shared during meetings was proof enough of the value they were adding to the lives of their students. Finally, value-added assessments are based on one test. When talking about the students within her own classroom, Loretta shared, “I’m not going to base their grade on one test,” [my emphasis].

**Teachers reported anecdotal gains in student achievement.**

With the exception of one, each teacher was able to describe gains in student achievement ranging from increased test scores to observations of student behaviors during the mid-evaluation check-ins. Although some might expect reports of student test data when speaking of student achievement, the strategic focus of the MUS teachers as well as their goals outlined in professional growth plans did not lend themselves to typical analysis of student test scores. As a school, two current site-level goals were to implement tenets of Visible Thinking and the Habits of Mind. Because of this, most participants’ goals were focused on developing skills in these areas rather than traditional academic subjects.

Despite the fact that goals were not academic in nature, participants still found evidence of progress towards their goals and impacts on students. Tracy, who was focused on bolstering students’ ability to recognize and identify Habits of Mind found that students did make gains. When studying figures in Black history, Tracy noted, “[Students] were able to find an important quality of that person and write it down and then find a sentence in the text to support that all by themselves. I was just really, really thrilled!” She went on to
describe how great an impact the use of a thinking routine, Three Words – Three Sentences, had,

I was not sure I could make it clear to them how to do three words, three sentences without demonstrating, you know, the same old way, but this way I think that they learned it so much quicker – more quickly.

She later shared that she felt students had exceeded her goal. She shared samples of students’ writing in which they clearly identified Habits of Mind that historical figures exhibited and supported those claims with evidence. Though she did not present any test scores, it was obvious that students had mastered the goal. Her feelings were echoed in her written reflection, where she stated,

As a result of the activities in studying Black History, the children seem to have a firm grasp of how to use the thinking skills of Same, Same, Different (compare and contrast) and 3 Words, 3 Sentences. They can identify character qualities in themselves and others and apply it to understanding characters in literature during reading workshop. The children spontaneously say, “that reminds me of..” when talking of characters in reading and writing or when discussing current events.

Nancy also incorporated more thinking routines into her instruction, yet she gave a more traditional assessment at the end of the unit. Although she did see gains in test scores, she was leery of attributing these gains solely to the increased use of thinking routines.

My kids did a lot better on the science test last time. So I don’t know if that’s because I put the thinking routines in or what happened . . . I did a lot more thinking routines at the end of my science unit on water . . . So, I’m not sure . . . but I know they’re understanding more.

Isabelle explained her take on student achievement similarly, noting increases in student test scores but having difficulty correlating that to the use of more thinking routines. However, she had students keep math journals and reflect on their thinking in an effort to gain insight into the reasons for the gains. Using the sentence starter “I used to
think . . . Now I think . . .” students were asked to describe the learning process in their own words. Though she did not have a full analysis available at this check-in, the method seemed an appropriate manner to measure student achievement.

Loretta had not given an end of unit assessment at the mid-evaluation check-in, but did note that she felt students were paying more attention to detail, using more academic vocabulary and demonstrating increased comprehension of word problems in math because of her increased use of thinking routines.

Cathy also felt there was evidence of growth in her kindergarteners’ achievement based on her observations of students. She was working on students’ persistence (A Habit of Mind) in writing and developed the tag line “stick with it” after reading the children a story on persistence. “The kids started seeing persistence in everything . . . kids were taking on this challenge in a new way.” She went on to say that she overhead students going back to this same language days later, saying, “I’m gonna stick with it . . . I’m gonna get it done.” She also observed a change in students work output. She used the phrase “stick with it” to encourage students to go back and edit their writing using a rubric. After this instruction, she had the students write spontaneously about a recent field trip, then edit their work. In her written reflection, she shared,

Almost every students’ writing had spaces between all of their words and they had hit almost all of the points on our collaboratively made writing checklist. This is a BIG change from what I was seeing previously. The students worked on their writing independently and I didn’t give them feedback as they were working. I did this to check in and see where they were at. It was clear to me after looking at their writing that it had improved.

All but one of her students included spaces in his or her writing compared to less than half previously. Perhaps more importantly, students were not only willing, but were excited to go back and edit their work, which they had not done before.
Although most believed students made growth at this point in the study, one did not. Nathan was hoping to increase his fourth grade students’ fluency in addition and subtraction by charting students’ progress on timed tests, setting goals, and reinforcing mastery through games. Without understanding why, his students’ averages steadily diminished rather than increased over four tests. He was quite perplexed with why this would happen. “I’ve done some pretty clear things in order to attain what I thought would be a relatively quick turn around [in fact fluency] and I don’t see that. So now I’m thinking, should I continue on in this form?” Although some might see this as a failure or weakness of the evaluation system, it in fact is not. Because this process caused him to reflect and look for evidence of learning, he was able to identify that his efforts were not working and he should try something else. Without the heightened attention on evidence, he may have just continued on in this same vein. Instead, he was forced to pause and reflect and had the added benefit of support from his peers to ask questions and give suggestions.

_Gains in student achievement, performance, and thinking were celebrated in final portfolio meetings and reflections._

Each of the seven participants who previously noted gains in student achievement continued to acknowledge their students’ growth in the portfolio sharing meeting and final reflections and interviews. However, many of the gains seen were not of the typical end-of-unit multiple-choice test variety. Teachers observed gains in motivation, persistence, use of academic vocabulary, conceptual understanding, and more.

Sarah sought to better differentiate her instruction in strings, teaching violin to second and third grade students. Because she only sees her students once per week, and sees over one hundred students, she needed a better way to individualize instruction. She
began implementing a data collection sheet where she could keep track of the specific focus area for each child. With this new tool, she had a much more systematic method to teaching students individually, and also had a record of past interventions to keep her organized and prompt her to follow up on specific skills with students week-to-week. At the portfolio-sharing meeting, she clearly demonstrated student’s improvement on specific skills by sharing before and after pictures and videos of students’ techniques, such as their bow-hold or posture. When reflecting on her approach, Sarah shared,

   It's just a small thing to implement but it made a big difference just for me as a teacher to not just say in passing like ‘fix your bow hold.’ It was like each week I needed to really check in and say, ‘okay how’s your bow hold this week?’

Cathy also used pictures and video to demonstrate growth in her students. She told the story of one student in particular who did not see himself as a learner. John, a young kindergarten student, was observed by many at the school to have a defeatist disposition. The art and music teachers, school librarian, and playground supervisors had all noted how easily John would give up and say, “I can’t do it.” As noted earlier, Cathy had really focused on persistence and her class termed the phrase “stick with it” as an agreement to how they approached their work. In addition to sharing writing samples, Cathy’s most poignant evidence was a video of young John after he completed writing numbers to 100, a task he would not have even attempted weeks before. In fact, this particular task was an optional challenge that students could do during their free time. John was smiling ear-to-ear in the video and the pride in his voice was evident as he explained that he stuck with it and did not give up.

   Cathy: John what are you doing?

   JOHN: I did the challenge.
Cathy: How did you do the challenge?

JOHN: Um.

Cathy: What’d you have to do?

JOHN: I stink-I stuck-I keep sticking with it.

Cathy also had evidence of more “academic” growth in her students. A second focus of her professional growth plan was to help her kindergarten students with their writing and transferring this persistence to editing their own writing and writing carefully with spaces in between words. By sharing pictures of student writing, it was clear to the study group that students had made tremendous progress in six short weeks. Below is one specific example of a student’s writing. In the first picture, the letters actually had to be cut and rearranged for them to be in any sort of readable format. The growth is evident when you look at that same student’s writing in the second picture.

Figure 4.2: Student Writing Sample
Cathy also shared video of her students using a rubric to edit their work, dictation of students talking about how persistence helps them be successful in reading and writing, and examples of students identifying characters demonstrating persistence in fictional stories.

Tracy had a similar goal with her second grade students and saw similar success. She hoped to build students awareness of the Habits of Mind as well as the role these habits can play in helping people find success. Through studying Black History, Tracy shared students’ writing samples in which they identified these habits in historical figures and compared and contrasted their habits with the people they studied. Nancy did the same with her sixth grade students and saw growth her students demonstrated through their writing.

Isabelle, Loretta, and Ethan all saw growth in their students’ metacognitive abilities. By using thinking routines in mathematics, each teacher collected writing samples, which
demonstrated students’ deep conceptual understanding of their respective topics of study. Isabelle shared, “I have expanded and pushed my math students to make their thinking “visible” and be able to give an in depth explanation about various processes and algorithms.” She went on in her reflection to explain,

> It is so rewarding to have a math class that commonly uses terms such as inverse, factor, product, divisor, dividend and quotient. Not only can they use the key vocabulary, but they can use the thinking routine “Same/Same/Different” or “Connect, Extend, Challenge” (and others) to show themselves, their peers, and the teacher the depth at which they understand each operation.

Nathan was the one candidate who did not report growth in student achievement, though his students did improve in their fact fluency on the final test they were given. However, as noted earlier, he grew as a teacher throughout this process and learned that his particular approach was not working. By trying a different approach, he did see gains in the final assessment of fluency.

It is obvious from these examples that teachers clearly felt their students benefited from the teachers’ personal growth throughout the evaluation pilot. Though there was no standardized state test data to analyze, teachers were confident that there students made gains in academics and learning skills.

> Participants felt current standardized assessments should play no part in teacher evaluation but a test more focused on critical thinking and problem solving would be an improvement.

Of the eight participants in this study, seven felt that current standardized test scores should play no part in an evaluation process. The one candidate who was not completely opposed to their use was the music teacher whose practice is not tied to standardized tests. Teachers felt the current tests are too flawed and do not measure what
is most important in education children. Tracy explained, "A standardized test is a one-shot inventory of what a given student remembers at a given time on a given day. It does not represent a consistent pattern of behavior for students." In her mind, a critical task of a teacher is to develop strong habits in students. Teachers are responsible for developing positive attitudes, dispositions, and more. Critical thinking was mentioned by six participants as an important aspect that is not currently tested well. Loretta shared, "They generally measure recall." Nathan elaborated,

Such tests basically measure the ability to respond to superficial prompts that do not encourage or even require complex, real world, solution oriented thought. Their emphasis promotes a sort of narrow-gauge curriculum and discourages imaginative and diverse thinking, often assuming that all test-takers have been exposed to identical intellectual, linguistic, and cultural backgrounds. These traits represent the antithesis of the direction that most teachers assert education needs to be heading in the 21st century. To disproportionately judge teacher performance based upon such test results can serve to dissuade educators from their valuable roles as intellectual risk-takers, deep and empathetic thinkers, and societal forecasters.

The group clearly agreed on the low-level nature of thinking required to perform on current standardized tests in California. In addition to this flaw in the tests, participants felt any test would suffer from the vast number of uncontrollable influences that impact students performance on assessments. Tracy posited that even, "the perfect standardized assessment as a one-shot inventory of student response [sic] will always be flawed." She went on to say that the test "would have to include many other factors including health factors, family influences, experiences, general attitude and skill in responding within a standardized test format." Though some might argue that a growth model would help account for some of these factors, Cathy disagreed. Having worked in both high income and poverty-stricken schools, she explained that getting a year's worth of growth out of students who are ready to learn is significantly easier than getting the same growth out of a
student with risk factors such as poverty, language barriers, or a disruptive home life.

Nancy agreed, stating,

If the child was several years below grade level when they entered the grade level, it would really be a challenge to bring that child up even one full grade level considering that there would be lots of kids on several different levels, so individualizing might be next to impossible and probably slightly ineffective.

Participants in this study struggled to imagine any fair system of using a standardized assessment to evaluate them. Even when given permission to have a “magic wand” to create an assessment that could reliably measure anything they could imagine, they struggled to identify specifically what should be measured. As mentioned earlier, critical thinking was the most common response. Five participants mentioned dispositions or habits such as persistence, empathy, emotional intelligence or the ability to deal with frustration. Tracy thought it would be interesting to measure students’ enthusiasm about learning as well as their independent work habits. Cathy felt it would interesting if, “there was a way to include a piece on habits of the mind as well since these habits are so important in all areas of a student’s life, not only academics. An assessment should evaluate what we want our kids to come away from school with.” Nancy suggested a potential question to illustrate the importance of learning strategies for students.

Students could read a passage about a situation in which there could be various solutions, and respond to questions about what strategies the characters used successfully or unsuccessfully to solve the problem. This would help assess the student’s knowledge of successful learning strategies. This would also force teachers to teach these strategies and reinforce them continually, therefore, teaching students how to learn using successful strategies. A KEY component to learning.

Isabelle suggested a similar scenario in which a social dilemma would be posed to students and they would be asked to respond from the viewpoints of various stakeholders in the dilemma. Nathan hoped to throw away the entire notion of normed assessment,
saying, “I think the essential task that any teacher performs is helping individual students progress. For this reason, I believe that teacher evaluations should be based on assortments of evidence that demonstrate development and improvement of individual students, not measured against a norm.”

Participants felt that even if standardized assessments measured what they view as important, the scores still should not account for most of a teacher’s evaluation.

For the reasons stated earlier, participants felt strongly that even valid and reliable test scores measuring the appropriate items should not be the most important factor in evaluation teachers. When asked what percent of the evaluation it should count for if the state said it must be included, participants responded with 0%, 0%, <5%, 20%, 30%, 33%, and 50%. Even when presented with the possibility of a “perfect” assessment, two teachers still felt it had no place in a teacher evaluation. They felt that the use of teacher portfolios and individual stories of student learning would be more appropriate. Cathy explained,

“I think that the biggest part of the evaluation should be about the portfolio because that represents more of what the teacher accomplished with their students. It also represents more of how the teacher is as a whole because it can cover more aspects of teaching instead of a snapshot on a given day. Including student work, observations, lesson plans, etc. really paints a clear picture of how that person is as a teacher. It also allows you to see teacher’s growth and thinking that they have made throughout the evaluation process, which is something that a standardized assessment can’t show.

Nathan added, “Unfortunately, for a variety of reasons, there are instances in which individual students may not be able to show progress despite a teacher’s best, and even outstanding, efforts. For this reason it would seem inequitable for educators to be evaluated solely on the basis of student progress.” From these statements and more, it was
clear that teachers felt their evaluation should rely on much more than test data, even if those tests were to improve.

**Teachers primarily view themselves as professionals**

As noted in the literature review, one premise of this study was that for a teacher evaluation system to be effective, it must be aligned with what teachers believe about teaching. During initial interviews (Stage 1), six of the eight teachers interviewed said they identified themselves as primarily operating in the professional realm of Darling-Hammond, Wise and Pease’s (1983) continuum. One identified herself as practicing craft and the other art. When asked about their peers, all eight described them as operating in profession, art, or a combination of the two. In teaching as a profession, teachers are expected to have prerequisite skills, but also to exercise judgment in their application. Teachers need theoretical knowledge as well as practical techniques and evaluations can focus on their ability to problem-solve. However, In teaching as art, teaching techniques and the ways they are applied may be novel and unconventional. There is a shift away from standardization to a process where teachers have to incorporate their pedagogical knowledge with their own personal resources and intuition. In this conceptual framework, evaluations might involve self-assessment and critical assessment by multiple people.

There was evidence that for some, the idea of teaching as a craft had a negative connotation. Though four participants said that they operate in the craft realm at times, where teachers are expected to have techniques with generalized rules for application. Tracy referred to craft as “narrow-minded.” When Isabelle was asked how the staff feels

---

6 Professionals are understood in the context of Darling-Hammond, Wise and Pease’s (1983) continuum, where teachers are expected to have prerequisite skills, but are valued for exercising their judgment. In this paradigm, teachers need theoretical knowledge as well as practical techniques.
about teaching as a craft or labor, she responded, “if you’re in those two categories then it’s … it’s not viewed in a positive light.” When speaking of teachers operating in the craft realm, Nancy stated, “They don’t tend to last too long here. They just have one way of looking at things. It’s like, ‘if the child can’t add you give him twenty problems instead of ten.’”

Another common sentiment expressed by four of the participants was that teachers should not always strive to be artists, creating new and novel experiences for students. For example, Sarah stated, “it’s always good to try new things and be innovative but not at the expense of skill and good judgment.” Loretta remarked, “I’ve seen some over time that treat it as an art and they’re not always successful because they’re too into the art category.” Isabelle added, “I wouldn’t feel comfortable with teaching something if it was always continuously new. There’s some sort of comfort associated with knowing what I’ve done before and feeling successful at it.”

Knowing teachers’ beliefs about teaching is important because of my theory that they will be more likely to be satisfied with an evaluation system that aligns with their beliefs. If teachers believed in teaching as a labor or craft, they would be evaluated via checklists and prescriptive criteria. The structure and language of the evaluation systems and forms would emphasize single-loop learning, where the evaluator scans the environment (observes), compares the information against operating norms (teaching standards and accepted practices), and initiates appropriate action (grades on evaluation forms) (Kauffman, 1980). This system does not encourage teachers and principals to question whether or not the operating norms, or standards, are appropriate. It simply measures the observation notes against a predetermined list of norms.
Knowing that most teachers consider themselves to be professionals and artists informed our later discussion of methods and structures that would support learning and growth. This belief that teachers are professionals did not change throughout the course of the research. In final interviews, each participant was asked to describe their beliefs and all agreed that teachers should primarily operate in the professional realm, exercising their judgment. Furthermore, as will be discussed later, teachers felt very satisfied with the pilot, noting that it treated them like professionals.

**The primary purpose of an evaluation system should be teacher growth.**

In addition to knowing what teachers believe about teaching, it was also important to understand what they felt the purpose of an evaluation system should be. In the initial interviews, each of the eight participants mentioned growth, professional development, improvement, or achieving goals as the reasons for engaging in an evaluation. When reflecting, Nathan described his past experiences with evaluations saying, “They had seemed like they were things needed to get checked off a list and moved beyond.” Cathy agreed, recalling, “sometimes it just felt like you’re doing it [evaluation] to do it type of a deal, versus getting a lot out of it.” These experiences left them yearning for a meaningful interaction with an evaluation system. Nathan described the ideal evaluation system as “a kind of machine, an engine for continued improvement, and continued motivation.” Tracy agreed and spoke to the importance of language, stating, “if you want to help me learn and grow, you, we, I wouldn’t call it an evaluation. I would call it coaching. Would you like someone to coach you to be better at your job? Yeah I’d love it. Would I like someone come in tell me if I’m doing a good job or not? Not really.” Research by Moss Kanter (1983) supports this conclusion, saying, “We need to create conditions, even inside large
organizations, that make it possible for individuals to get the power to experiment, to create, to develop, to test – to innovate” (1983, p. 15).

The belief that the purpose of teacher evaluation is to help teachers learn and grow was reiterated in final interviews. All eight participants agreed that this belief had not changed. However, six of the eight also included motivation as a piece of the purpose of evaluation. They felt that evaluations should motivate teachers to try new things, reflect on their teaching, and engage more deeply in what they do, leading to growth. Nathan explained,

I’ve never read anything that suggested that teacher growth, teacher competency, was not the essential factor in educating children. So, it kind of seems like without a lot of fanfare, or further education, we would be able to say, anything that we can establish that promotes teacher growth gets us directly to our overall goal, which is student growth and achievement.

Nancy, who plans to retire at the end of the year, added, “If you’re going to have 43 years in it [teaching], there has to be something in it for you; it has to be learning for you and from learning comes excitement, and change.”

**Participants felt that a focus on teacher growth would lead to improved test scores.**

The teachers in this study felt that overly focusing on student test scores was like a doctor treating the symptoms, not the root cause of an issue. Judging teachers by one end-of-year test is almost like an autopsy, attempting to help a teacher after the work of teaching has already been done. How much more powerful would it be to analyze teaching while it is happening and focus on growth? Nathan demonstrated his thinking through a classroom example, saying,

Maybe they’re not scoring on those science tests but let me show you some pictures of these kids doing science... here’s something that we’re getting at, and it may be
one sentence that’s poorly constructed but it’s showing [that] the kid’s getting the idea. You know, you can see that there’s the science going on in the kid’s brain.

An end of the year science test might tell a teacher how their students did on the test, but the use of portfolios can help a teacher see how students are doing science. Furthermore, much more information is gained from seeing how a student thinks about a scientific concept than from whether or not they answered a multiple-choice questions correctly. Isabelle shared, “If the teachers are growing professionally and making professional goals, then, in turn, the students are going to have higher achievement.” Nancy added, “You need to look at the teacher and see how dynamic the person is, how they change over time, because those are the people that are going to be teaching dynamically.” Isabelle’s simple statement, “Good teaching is good for the kids,” summarized the group’s beliefs. An evaluation system that helps teachers develop good teaching will have great impacts on student achievement.

**Systems of evaluation must develop teacher’s ownership of the process and motivate them to engage in thinking.**

When asked what an evaluation should do well in initial interviews, Nathan’s first response was, “Well, the first one is - motivate teachers.” Nancy described intrinsic motivation when given the same prompt. Loretta’s first response to the question was “motivation – an evaluation would do well if it motivates and inspires a teacher to want to improve their craft, to want to meet their students’ needs better.” Ethan’s initial response was that an evaluation should inspire and engage. These teachers were all looking to the evaluation as a tool to reinvigorate them and move them towards what they saw as the purpose of teacher evaluation, growth. Ethan summarized, saying that the conversations
that would likely occur in an evaluation system should be a “catalyst for people to be engaged.”

Knowing that teachers want to be motivated, how they are motivated needs to be understood. Nancy expressed how motivating a goal setting process was to her. She recalled goal setting with a previous administrator.

“Rather than someone else telling you, you have to do this, then the teacher would figure out what they had to do. It gives you power. And it, and makes you feel proud at, when you come in and you have it all, you’re kind of giddy. It’s pretty fun!”

One can sense the pride in this statement at having grown and taken ownership for the process. Three other participants expressed this same sentiment. Involvement and ownership in the process through goal setting would be motivating to them. This aligns with Fairtlough’s (2007) suggestion to an alternative to the traditional hierarchical structure of most schools. In a responsible autonomy, individuals or groups have autonomy to decide what to do, but are accountable for their decisions. The goal setting process and collection of evidence is a means of providing some autonomy while holding teachers accountable.

For two of the teachers, the mere fact that someone else is in the room and watching them teach would be motivating to them. Schein (1995) describes why this and other methods of evaluation such as peer observation and the use of data can lead to change. Change begins with disconfirmation, or dissatisfaction and frustration that arise from data that do not meet one’s expectations and hopes. Dissatisfaction with one’s own teaching might be induced by observing other teachers who one sees as superior to his or her own teaching or by feeling like someone witnessed their own sub-par instruction. However, disconfirmation alone is not enough to change behavior. Some may ignore the information
or blame it on external factors. In order to accept the disconfirming evidence, it must connect to something that the teacher cares about, inducing a feeling that not changing will lead to failure of some sort. Schein refers to this feeling as *survival anxiety*. By accepting our failures, we are in danger of losing effectiveness and losing face with our peers. Though not explicitly addressed in the interviews, these concepts are important to consider in the design of an evaluation system.

Each participant in this study clearly expressed how they found the process motivating in the final group meeting and individual interviews. Tracy described the system as “invigorating,” explaining how she was excited to think about her teaching and anxious for the next opportunity to discuss what she was doing with others. Ethan described the process by explaining, “It’s not a job that you’re just here doing. It’s a profession in which you can be growing and thinking and you’re intellect is connected to what you’re doing.” Nathan commented that he was also motivated by seeing the work others were doing. To be exposed to the ways in which his colleagues thought about their practice sparked him to think about his own practice in new ways. Nancy shared how her sense of pride and accomplishment was motivating, saying, “I think that’s very empowering. You feel like a better teacher. So then once you start feeling like a better teacher you start striving for even more lofty goals.” There was an exponential effect to her success. By feeling successful with her original goals, she was motivated to take her practice even farther. Tracy, who plans to retire at the end of the year, compared this process to a more traditional process, saying that in the past, “It wasn’t a learning experience it was just passing a test. This, I mean, really, I was excited! That’s pretty good for an old fart.” She later added, “I found it joyful. I mean it was just… it made my work
exciting.” Loretta added, “It’s inspiring! It’s motivating!” Teachers were unanimous in agreeing that the system motivated them to improve their practice.

**Teachers want to engage in this growing process in a safe and supportive environment.**

According to MacGregor’s Theory Y (Northouse, 2007), the hierarchy between administrator and teacher needs to be flattened, allowing for trust and cooperation to occur between the two. This idea of safety was present at multiple points in the research. In the initial interviews, seven of the eight teachers agreed that feeling safe and supported is critical when engaging in a growth process. Nathan noted that without trust, the evaluation process becomes a “shell game” where teachers put on a show, trying to demonstrate what they think the evaluator wants to see. Tracy concurred, sharing that an evaluation, “Should not make you feel like crap. It should not be a threatening thing. It should not be, ‘I’m going to come in and see if you could read my mind and teach the way I want you to teach.’” The concept of honesty appeared in numerous answers as well, with teachers stating that an environment should exist where teachers can be truly honest about their strengths and weaknesses and expose themselves to others. When asked to name two things an evaluation should do, Isabelle said it should be, “open and honest” and that she wanted to “feel supported by my administrator.”

**The role of the administrator plays an important part in creating a safe and supportive environment.**

In the final group meeting and individual interviews, participants explained how my administrative role as a facilitator helped create a climate of safety and trust that supported and encouraged growth in a non-threatening way. I intentionally attempted to
remain neutral throughout the process and not insert my own personal judgments about the individuals’ teaching and thinking. Instead, I took the stance of being there to support teachers in any way I could by connecting them to resources and asking questions. Ethan remarked,

It’s not about having a structure that you’re holding people to. It’s about providing an environment in which people feel empowered and passionate about their work. And supporting them to be better people, to be more passionate people, to be more interested in what they’re doing and to give them the tools and the time to go out and do things well. And have an environment in which they feel supported.

He went on to say the administrative role is about, “helping people think about their process and helping them think about what they’re doing, and where they’re going.” Loretta likened it to the culture she tries to create when teaching her fourth graders to write. She tries to create a safe climate where they try new strategies and styles. She exposes them to new ways of thinking on occasion without telling them how she wants them to do it. She would like an administrator to do the same thing with teachers; give them ideas without judging. However, she also explained how she provides guidance individually and in a gentle and reassuring manner and would like an administrator to act similarly with a teacher, saying, “Hopefully it [corrective feedback] would be done one-on-one as opposed to in front of everyone.” Tracy also appreciated the lack of judgment from the administrator, noticing,

I felt good that you were so nonjudgmental. I mean, you probably had thoughts in your head, but, they didn’t come out of your mouth. And it was just, it was very supportive. I didn’t feel like I was being judged. I felt like, oh, if I needed a resource you would help me find it.

She echoed these sentiments in the initial planning stage as well when she said,
“If you want to help me learn and grow, you, we, I wouldn’t call it an evaluation. I would call it coaching. Would you like someone to coach you to be better at your job? Yeah I’d love it. Would I like someone come in tell me if I’m doing a good job or not? Not really.”

Nathan saw an opportunity use the administrative role to build upon Schein’s (1995) stage of disconfirmation, believing, “the administrator’s role is to encourage and acknowledge, and to celebrate moments of disconfirmation.” By highlighting moments when a teacher comes to new understandings, the teacher better recognizes an opportunity to improve his or her practice. The administrator may not explicitly judge the good and the bad in a teacher’s practice, but can shape how a teacher changes by highlighting those instances of disconfirmation. However, an administrator may not have to judge at all. Nancy explained, “You don’t have to be judgmental because I think the group will provide that.” The use of study groups provides natural opportunities to peers to explicitly or informally evaluate one another’s work. (more regarding peer influence on page 102).

Although I made every effort to articulate my role as a support and facilitator, Isabelle felt that it would be impossible to completely eliminate the hierarchical nature of the administrator-teacher relationship. She shared, “Even though you didn’t come in on a top down theory, so to speak, you’re still an administrator. You’re still essentially, on some level, observing and taking notes on what we’re doing. I think that element is still there even if it’s not obvious.” Sarah agreed, saying, “When you came to observe, I still felt like, okay, my boss is in the room, you know.” It may be important to note that both of these teachers were on the relatively inexperienced end of the spectrum of participants in this study. Despite this unspoken reality, all eight participants did vocalize that they felt
comfortable, safe, and supported throughout the process. Reasons for these feelings will be further discussed in Chapter 5.

Participants were excited about the use of videotape and peer observation as evaluation tools.

The fact that participants were interested in videotape and peer observation served as evidence of their feelings of comfort and safety. These methods had not been used at this site in the past, yet the research suggested they were valuable methods and the participants had demonstrated a preliminary interest in them during the initial interviews and group meetings. Five participants listed videotape as a piece of evidence and six listed peer observation, with only one participant, Nathan, not choosing either. This could be considered an indicator of psychological safety, as one might traditionally expect teachers to be nervous about the scrutiny of their peers and the permanence of being videotaped. Their willingness to be videotaped is important to note, as Marshall (2005) notes that watching a videotape of one’s lesson with a critical friend is one of the most powerful ways to grow as a professional.

It is also important to note that although Cathy was one of the two not to choose peer observation, she also did not have a grade level or content-focus peer in this study. For the other six, there were natural partners to engage in peer observation with. Isabelle, Loretta and Ethan were each interested in working on thinking routines in mathematics, so it would make sense to watch and learn from one another. Tracy and Nancy were both engaged in the study of Habits of Mind through Black History Month and even planned to co-teach some lessons together. Although Sarah did not have a music peer in the study group, she had an existing relationship with another music teacher and had already
planned on doing some peer observation prior to this study. Cathy’s plan’s absence of peer observation may have had more to do with lacking an appropriate peer than disinterest in the method itself.

Another interesting observation to note is that four of the five participants agreeing to videotape lessons were in Group 1. Early on in Group 1’s discussion, Ethan shared an experience a colleague had in which the use of video had greatly enhanced his teaching. This lead to a dialogue about the use of video, participants’ experiences being videotaped, and the merits of it. The conversation seemed to excite people about the prospect of using video as a tool for learning and growth. In Group 2, the only teacher to use video was Cathy, and she noted that she was already videotaping conferences with students, so was comfortable with the idea. She was also the second to last to share her plan, so the idea of using video was not introduced until later in the session. A peer did suggest video to Sarah later, who presented last, but she ultimately did not include it in her plan.

**Participants were willing to demonstrate vulnerability through admitting a lack of knowledge or certainty.**

During development of professional growth plans, the phrase “I don’t know” was used by six of the participants when discussing their professional growth plans. The teachers felt safe to admit a lack of knowledge or conviction on what they should do and how they should do it. When trying to set a realistic goal for mathematical fluency, Nathan admitted, “I have thought about it and I don’t have the answer, I mean, I don’t know … I think that’s a good question. I don’t have the answer to it.” When talking about her students’ writing development, Cathy wondered aloud, “I don’t know if it’s appropriate right now to say they need to be accurate.” Sarah shared her struggles with music
instruction, saying, "I haven’t figured out how to balance whole versus individual very well.”

Participants continued to feel comfortable expressing uncertainty and asking for help in the final group reflection and individual interviews. Nathan stated,

I kind of felt like the environment is such that, to any of those people, I could go to them and say, “How do you do that? Show me more about that. Tell me. I don’t know about that.” And that’s not going to hinder my reputation or my progress.

These anecdotes and more point to participant’s willingness to admit that they were still learning and growing and lacking in knowledge in front of both peers and an administrator.

Participants demonstrated openness to suggestions for means of collecting evidence.

Throughout the pilot implementation in Stage 4, all participants demonstrated openness to new ways of collecting evidence. This could also serve as evidence that teachers felt safe to give and receive feedback. This readiness could have been heightened because the process of collecting evidence was a novel practice for all of them. After presenting her goals, I asked Nancy if she would be willing to be videotaped and she responded, “Oh, sure. Yeah, I think that’d be really fun.” It is important to note that as the administrator, I was not the only one offering suggestions. In fact, through explicit questions, I was able to engage others in the process. When Cathy struggled to think of a means of measuring whether or not requiring accuracy and precision in kindergarteners writing might inhibit their love for writing, I asked, “Can you or anyone at the table think of a way to document how it’s hindering or helping a kid’s love of writing at all?” Nancy responded,

Well, you could ask them to, you know, if you got the first piece that they wrote and they’ve got the next piece and then you say, “Which one do you like better and
why?” You could do a videotape, or, whatever. Uh, then, [ask] ‘Which one do you like better’ [which] is probably the same as ‘Do you like to write’. Right? Because, when kids take pride in their work, they like it.

On a number of occasions, participants included evidence offered by their peers, but I also found myself offering many suggestions for the staff that centered on work I knew they already engaged in. For example, I knew that Nancy had a planning day scheduled with a peer that focused on items related to her goal. By suggesting that she keep the day’s agenda, notes and work products, she would have strong evidence without requiring “extra” work. Knowing that each grade level engages in weekly meetings with focused conversations, suggesting that teachers provide meeting summaries was another simple piece of evidence they could provide. I suggested peer observation to Isabelle, knowing that was something her grade level already practiced.

*Participants felt psychological safety at the mid-evaluation check-in.*

At the mid-evaluation check-in, participants noted how peer interaction specifically helped build their sense of safety and freedom to take risks. Isabelle shared,

There’s definitely some comfort knowing that we’re in a group together so I can talk to my colleagues and say ‘What do you think about this? Is this working for you? Do you feel the same?’ or just touch base with them. So from my standpoint I feel like it made me grow much more than it would if I just had three formal evaluations.

Paradoxically, some participants felt comfort in knowing others were uncomfortable, just like them. Isabelle sought comfort in knowing someone else was nervous about implementing thinking routines.

Isabelle: How do you feel about it? I mean do you feel comfortable? Because I felt initially uncomfortable with the process and then it got . . .

Tracy: I felt totally - it totally freaked me out to see how effective [sic]. I was not sure I could make it clear to them how to do three words, three sentences.
Knowing that they were not alone made them feel better about this work. It made it acceptable not to be perfect, and therefore it made it acceptable to take risks. Tracy continued, saying how different this was from a typical formal observation by an administrator, explaining, “If I knew that you were going to come in to observe the [one] lesson I would plan something I knew I would be successful at and not take any risk. And, this was more exploratory and I – I learned a lot more from it.” This is important to note when thinking about creating a safe environment for learning.

Besides the involvement of peers in an evaluation process, participants noted that the ongoing and frequent nature of the evaluation also lead to safety because they knew the evaluation did not rely on a single observation. Isabelle supported this statement when she said,

I think it makes it a little less anxious for a probationary [non-tenured] teacher because you know that you’re [the administrator] going to be visiting often so it becomes more of an informal process rather than one formal hour, three formal hours throughout the year that you get to see me. And plus you get to get a better feel of really what’s happening in my room 9:15 to 10:15 rather than a snapshot. So that’s felt good and I feel like as a teacher it allows me to grow a little bit more in taking a responsible risk.

A system of evaluation needs to be flexible and responsive to the varying needs of diverse teachers.

Though one might perceive how a flexible system that responds to teachers’ needs could also promote safety, the importance of this facet of the system deserves its own section. Teachers at MUS did not believe an inflexible tool would support teacher growth and development. After reading the literature review, Isabelle noted, “It is apparent that a one-size-fits-all model doesn’t seem to be effective in improving teacher practice.” This same sentiment was echoed by two other respondents. Loretta mentioned how a multi-
faceted approach could capture more of the essence of teaching as well as their learning throughout the process. This aligns closely with teaching as a profession and art. In teaching as a profession, Darling-Hammond, Wise and Pease (1983) suggest evaluating teachers ability to problem solve and in art, evaluation involves self-assessment and critical assessment by multiple people. Both of these goals could be reached through using a multi-faceted approach such as a portfolio.

*The participants were interested in a flexible system and saw goal-setting and the use of portfolios as a means of achieving this goal.*

There was clear consensus that an evaluation system should be flexible. Tracy used the analogy of students learning in different ways to argue in favor of a dynamic system for teachers.

In your classroom, children don’t all learn in the same way and you, and I’m thinking that teachers don’t always all learn in the same way either... for me the best world would be a menu of options for how you want to continue your professional growth.

The group quickly responded to this notion of a menu of activities and evidence and demonstrated excitement to explore the idea further, with responses and suggestions being offered in rapid succession. Noting that the menu gave them more ownership and control of the evaluation, teachers agreed that a menu or portfolio would help them grow as teachers. However, the group also noted a needed for some clarity and guidance. When considering a completely open and flexible system, Nancy expressed, “Frankly, I would need some more boundaries than that.” Cathy suggested a system often used in primary classrooms, employing a “must do” and “may do” list, in which key elements like observations and reflections must be done by all, but also allowing teachers to choose from a list of “may do’s.” Adding even more flexibility to the system, the group agreed to
Nathan’s idea of including a “defensible other” as a menu item in which a teacher could propose any item not on the list if they could defend why it was appropriate.

Participants continued to value the flexibility of the system in the final group reflections and individual interviews. Sarah posited, “It’s about you ... It’s your growth ... even if it’s two second grade teachers, you’re different people. You grow different ways.” They felt the flexible nature of goal setting and evidence collection supported their individual growth. Nancy reiterated how the portfolio process itself was a great example of flexibility. Some participants came with binders while others had slideshows or PowerPoint presentations. Both methods were capable of demonstrating growth towards one’s goals. Isabelle added, “There’s such a variety in how you want it [the portfolio] to look like or what you want it to look like, then that gives you some ownership in it.” She went on to explain that if we had one model, teachers would create what the administrator wants, not what would be personally valuable to them. The flexibility allowed them to engage in the process and do it for their own personal growth.

Despite the value of flexibility though, participants felt there should be some guideline or universal truths. Much like the conception of teaching as a profession, we generally know what works, but need to exercise judgment in applying these concepts.

Nathan explained,

I think any institution or cohort is going to have a slightly different approach, and I think that’s fine. But the core pieces of providing the teachers with information, providing a safe and collegial environment, and the continuous sharing of progress are the big parts.

There is a difference between a completely flexible system, and one in which flexibility is allowed within a certain construct or parameters. Each of the others agreed in the final interviews, asserting that items such as peer collaboration, evidence collection and goal
setting must be present, though their form may differ from teacher to teacher, or even at a system-wide level school to school.

**RQ3: What do Teachers at Mountainview Unified School Believe are Appropriate Ways to Evaluate Teachers?**

*Teacher growth is supported through various methods, including peer interaction, goal setting, evidence collection, reflection and an ongoing and iterative process.*

The discussion of beliefs above also led to insights into research question three, which is focused more on methods than beliefs. However, the two are closely interrelated. For example, a measure of guidance we agreed to was adding explicit goals and a plan of action in outlining portfolios\(^7\). Goals could serve the dual purpose of allowing for some teacher control in setting goals, but also some clarity and guidance for what the teachers were working on. The group felt that goals would help focus their work and make the growth process more explicit. Furthermore, they felt goals should be set in a pre-conference setting with the administrator to ensure that they were reasonable and appropriate. This specific method was seen as supporting their belief that a system should be flexible. The findings below summarize the participants’ beliefs and experiences with the methods and logistics of evaluation, though the methods are also rooted in their beliefs. Individual interviews, field observation notes and written reflections paint a picture of teachers’ beliefs about the logistics of an evaluation system. Teachers valued opportunities

\(^7\) For the sake of this research, “portfolios” are defined as any collection of evidence including but not limited to written reflections, student work samples, photos, video, agendas, meeting notes, conversation summaries, etc.
for peer collaboration, goal setting, evidence collection, reflection and an ongoing and iterative process.

**Teachers believe in peer observation and collaboration as appropriate and motivating facets of teacher evaluation.**

Throughout the study, the power of peers interacting with one another came up again and again. The system itself was designed to provide opportunities for peer interaction through the use of study groups, but participants desire for peer interaction was seen in their individual choices as well, choosing to team teach, observe one another, or plan together throughout the pilot implementation.

**Teachers expressed a desire for peer collaboration in initial planning.**

Six of eight participants spontaneously volunteered peer observation or collaboration with peers as a valuable tool in teacher evaluation. This supports Schein’s (1995) claim of the need for psychological safety and teachers’ own reports of wanting a trusting a supportive environment. The less hierarchical relationship between peers can make the process less threatening. Peer observation and collaboration can serve multiple purposes; creating psychological safety, providing disconfirming data when one compares his or her practices with peers, a sense of survival anxiety, or the idea that not changing will lead to failure, and finally cognitive redefinition. Ethan likened the process of cognitive redefinition that happens during peer observation to students copying one another in class.

The first six weeks you’re going to find that a lot of kids are just looking at what the kid next to him is doing, and doing the same thing, but after five, six weeks; all of a sudden they’re starting to do their own thing. And, I think that would happen with folks that are struggling, if they’re in a group.
Teachers can broaden their knowledge through cognitive redefinition (Schein, 1995). This includes learning new meanings of terms, broadened interpretations of concepts, and new anchors and scales of judging and evaluating. One way in which these new ways of learning are introduced is through imitation and identification with a role model. When peers engage in meaningful conversations and share their interpretations of lessons and teaching activities with one another, they can be motivated to change. For all of these reasons, the benefits of peer observation and collaboration are obvious. Sarah summed it up, saying, “I really think it would be valuable for our peers to evaluate one another.”

*Teachers reiterated their belief that peer interaction is an important piece of an evaluation in Stage 3.*

As noted in the initial interviews, participants revisited the benefit of learning through interactions with peers in the pilot planning session. This could occur through more formal structures such as observations or informal conversations that happen in passing. Based on the number of times peer interaction was brought up in our conversation (it was mentioned 19 times), it was obvious that this is something they valued. Nancy noted how valuable it could be just being exposed to different ideas. She also noted a desire to be proud of her work and share it with others. Sarah shared that it is comforting and motivating knowing that others are going through the same process and learning together. Later in the conversation, Loretta suggested the power of intentionally diversifying peer interactions by creating mixed groups with varying grade levels, experience or subject area. All agreed, and Ethan pointed out the power of hearing from someone else who might be looking at the same issue through a different lens.
Participants noted the value of working together in creating their professional growth plans.

Participants recognized the value of working together to clarify their plans for professional growth. When Isabelle shared a strategy she uses to organize thinking routines, the following dialogue took place:

Tracy: Oh that would be great, see how helpful it is to... [work together] [LAUGH]

Ethan: Yeah, see? [LAUGH] The power of team.

Tracy: Thank you!

Later, Loretta was informed that another colleague on campus had done some work very similar to what she hoped to accomplish, and she responded, “Oh my gosh! Then, I want to talk to [her] then.” When Tracy concluded her part of the discussion, she was still eager to continue to interact with peers, saying, “I feel good, I'll probably have questions [later] and want to pick someone's brain.”

Participants expressed the value and benefits of peer interaction at the mid-evaluation check-in.

When Nathan struggled to improve his students’ mathematical fluency, the process of interacting with a study group of his peers led to new thoughts and understandings. Through multiple rounds of questioning and offering advice, Nathan began to think about the issue in new ways and left the meeting with ideas to implement. His group suggested that the students might be suffering from anxiety when taking timed tests. Rather than expecting all students to complete the test in three minutes, perhaps he could just set a running clock and record students’ finishing times. That way, they might feel successful from the start and be motivated to beat their last time on subsequent tests. The exchange
below also highlights how Nathan was exposed to new thinking through interaction with peers.

Sarah: Yeah I was just thinking out loud. What’s the breakdown?

NB: Yeah. Huh.

Cathy: Yeah it might be interesting though, to talk with them and just say ‘you know, why do you think you’re not able to complete it?’ or you know like what ...

Nathan: Yeah ‘cause they see it now. They’ve charted it ...

Sarah: Yeah.

Nathan: ... and they know.

Sarah: Do you know a lot about how it’s being taught in the lower grades?

Nathan: No-no-no. I don’t know. Yeah I understand what you’re saying.

In this quick and simple exchange, Nathan was reminded to think about identifying where the breakdown in success was happening. He was given the suggestion to directly ask the students why they thought they were not being successful. Finally, he was indirectly given the idea of talking to the teachers in younger grades to gather more information on how they taught addition and subtraction.

Nathan was not the only participant to benefit from peer interaction as part of the evaluation process. Each participant benefited, and benefited in different manners, including exposure to new ideas, understanding their practices in new ways, fostering teamwork, easing anxiety, or simply being sparked to action. The most commonly reported benefit was the infusion of new ideas by listening to what others were doing or being offered suggestions. Nancy shared, “I think teamwork is really powerful, really powerful, and it’s been - that’s one of the best things that ever happened to me.” When Sarah heard that Cathy was using the phrase “stick with it” to encourage her students to be persistent,
She immediately thought about using that same language when she teaches music to Cathy’s students and other classes as well. Nathan made a new analogy between Sarah’s work with students and his own students struggles in addition and subtraction. When she spoke of her students struggling to overcome an improper bow-hold when playing violin, Nathan saw a parallel to his own students overcoming bad habits in addition and subtraction such as using fingers to count. Two professionals working in completely different circumstances saw value in thinking about teaching together. Nathan went on to compare the process to the medical field.

You gotta start like physicians. They start by eliminating, you know, what it is, what isn’t it. And so I think to break it down and try to break it down from a variety of perspectives helps me. So I feel like I’m on a hunt now.

Throughout these conversations participants were given suggestions, explained how another participant gave them an idea to try something new, or learned a lesson from someone else’s experiences.

Participants highlighted peer interaction as a critical component of any system in the final group reflection and individual interviews.

Though the power of peer interaction was noted throughout this process, peer interaction and collaboration were highlighted as one of the key ingredients in the success of this system when reflecting on the process as a whole. When asked what components of an evaluation system are most important, Tracy’s immediate response was, “I would say make sure you have the study groups, make sure that you keep them small enough that they can build trust.” The interchange below more fully describes what led her to that conclusion,

Tracy: I loved working with the other people of other grade levels. Especially having a mix of grade levels.
Q: What did you love about it?

Tracy: What I liked is that they didn’t ask the typical questions that you would get from someone in your own grade level.

Q: Because they were really trying to understand?

Tracy: Yeah, they were really trying to understand what you’re doing and they’re coming from another age group, so their questions were a little different and it made me think a little more deeply. Because I think sometimes, if it’s just always first grade talking to first grade, they don’t get that stimulation to bump it up a little or...

Q: Especially in a one-school school district where it’s not like you’re comparing yourself to teachers at other schools.

Tracy: Yeah. And it stays stagnant, sometimes.

Q: Did you feel safe?

Tracy: Yeah.

Q: Was it competitive?

Tracy: No, no, … and I had doubts going into it that I would… that it would work for me.

Q: What were you concerned about?

Tracy: I thought it would just be blah, blah, blah, waste of time. Because I’m so smart. [SARCASTIC] I actually learned a lot, and it was such an equal sharing with no pressure. It wasn’t like you’re expected to do anything except be yourself and participate as a professional. Which I loved.

Other teachers illuminated the unexpected learning that came about through peer interaction. Nathan referred to it as “serendipitous learning,” saying, “There’s all kinds of unanticipated learning that can take place.” By regularly interacting with peers, the learning extended far beyond the scope of original goals. Even Sarah, as a music teacher, felt the benefit of peer collaboration with classroom teachers. She stated,
Mark’s topic really didn’t apply to me at all, you know, but I found it really interesting to hear about his process and the way that it went and didn’t go the way that he expected it to. And also there was, you know, a level of – I found myself thinking.

In addition to learning from peers, five participants explained how peer interaction helped them stay accountable and raised the bar regarding expectations. Ethan analogized,

Sometimes the reason I’m not getting excellent work from my students is because they don’t know what excellent looks like. And that is true among teachers too. And opportunities to view excellence I think are very powerful. So part of this cycle again, is to make opportunities for teachers to see their colleagues at work.

Isabelle added, “The accountability element is much different because you have colleagues who are watching you, and discussing how you’re doing, and what you’re doing, and why it’s working, and why it’s not working.” Nancy agreed, saying, “Meeting with the groups along the way was great because you got to figure out what other people were doing and how they were coming, but mostly I think it was to hold yourself responsible.” Knowing that they would be expected to speak thoughtfully about their work and share evidence with one another motivated them to be better prepared, and therefore more engaged in the process. It is clear through these examples that the benefits of peer interaction were many.

**A system of evaluation should be ongoing and frequent.**

One weakness of the prior system of teacher evaluation at MUS and many other schools is the length of time between feedback cycles. With observations occurring only once every one to five years, this minimal feedback is highly unlikely to change behavior. In initial interviews, participants shared their beliefs about the frequency of evaluation. Loretta summarized many of the participants’ experiences, saying, “I don’t remember a lot of formal evaluations in recent years.” Cathy, Tracy, Sarah and Isabelle shared being frustrated because a single observation was just a snapshot of their teaching, and not a true
picture of their capabilities. Isabelle got to the root of the problem when she said, “essentially our evaluations as a teacher are all summative, and only just a little snapshot of what happens in our day.” For an evaluation to really be a growth opportunity, it needs to shift from a summative to a formative approach.

All eight participants mentioned that evaluations should be more frequent and/or ongoing. Nathan shared, “I think the regularity is important and it needs to be viewed as part of an ongoing process.” Ethan concurred with, “I just think it's an ongoing process everybody should be doing every year... almost like a piece of writing you know ... You can always improve it, you can always make it better.”

Interestingly, increased frequency could lead to more psychological safety. Nathan made this same connection, saying “the more regularly it [evaluation] occurs and the more ... familiar you become with a process the less anxiety it can produce.” Nancy noted, “It really feels good actually when an administrator really knows your teaching.” This thought was confirmed when she stated in her final interview, “I thought this was a lot more productive. I used to get really nervous for the one lesson a year. This was a continuation and I realized it was for me; it was for my growth as a teacher and not just to satisfy you.”

Cathy explained that with infrequent formal observations, things would happen before and after the actual observation that she often wished the administrator would have seen. With a more frequent and ongoing process, the administrator is much more likely to see these great things happening. When asked which system they felt provided a more accurate picture of one’s teaching, each of the eight participants agreed that the ongoing and more thorough nature of this system allowed for a fuller and more accurate picture of someone’s
actual teaching, building a sense of safety in knowing the evaluation would more likely be accurate.

Isabelle explained how the frequency held her more accountable and forced her to embed new ideas more thoroughly into her curriculum. “I’m taking that one goal and really putting into my everyday teaching; whereas the prior model of evaluations … I would do it three solid times while you’re watching me.” Because she knew I might drop in at any time during the math block, she was motivated to truly embed this new practice into her everyday teaching rather than only implementing it when she knew I would stop by for a formal evaluation. Cathy expanded on this idea, saying that in order to fully understand any new practice, a teacher has to implement it multiple times along the way and reflect on why it did or did not work, tweaking it along the way. Having a long-term process with ongoing monitoring provided teachers with the opportunity to experiment and get feedback along the way, increasing the likelihood of disconfirmation.

**Reflection is an important part of learning and growing.**

When a staff embraces teaching as an art, Darling-Hammond, Wise and Pease (1983) recommend self assessment and reflection as a tool for evaluation. Feeney (2007) agreed with its importance, stating that evaluations have no meaning unless there is reflection. Of the eight teachers included in this study, six noted the importance of reflection in their pre-interviews. Nancy said, “Reflecting, I think, is the critical thing,” and went so far as to say, “You should never hire anybody who can’t reflect.” An evaluation should help a teacher think about what they can do better and honestly evaluate their strengths and weaknesses. Reflection could also be an important component of broadening teachers’ knowledge through cognitive redefinition (Schein, 1995), or broadening
interpretations of concepts and developing new anchors and scales of judging and evaluating. The cycle of learning and growth is incomplete if participants do not come to new understandings. Reflecting is one means of articulating and solidifying one's understanding.

This belief in the value of reflection was specifically reiterated by five participants in the final interviews. Nancy was the biggest proponent of reflection, advocating that, “To be a great teacher, you have to reflect daily. Reflect on the lesson. Reflect on what’s gone on in your classroom that day... Reflect on each lesson even as you're teaching it.” Sarah added, “We can get so busy as teachers to where we don’t really take the time to reflect always, and it was this sort of forced reflection that I think is really valuable.” Isabelle explained how the use of portfolios caused her to reflect and provided opportunities for cognitive redefinition.

Putting something together, an end product to be able to go back on and say “this was my goal for this school year, for 2011 and 2012,” and to be able to just reflect on that as a professional, is a nice tool in itself. It takes time and it takes energy, which I’ve said before, but it's true. And then when you’re sitting down and you’re organizing all of these thoughts on where was I, where am I now, where do I want to go, it kind of solidifies and makes everything more concrete.

As noted below, Isabelle was not the only participant to find value in the use of portfolios.

**Teachers saw value in the use of portfolios as an evaluative device that provides evidence of teacher and student growth.**

Portfolios, or collections of evidence of a teacher’s work, were an idea that three out of five teachers wanted to utilize after reading the literature review. Cathy said, “I really connected with the idea of teacher portfolios as I read the literature review. Portfolios seem to be linked to the idea of teaching as a profession and as an art.” This is important to
note, considering the theory that an evaluation system should be based on teachers’ beliefs about their work, in this case remembering that seven of the eight participants shared that they primarily view themselves as operating in the realm of profession and/or art.

Portfolios also connect to the idea of growth and motivation through ownership of the process and peer collaboration. Cathy noted:

A portfolio at MUS would be a great idea because it would allow teachers to choose an area or areas to focus on and it would allow them to see their growth in these areas over the years. It could also serve as a way for teachers to collaborate if they were focused on similar areas. It could also give teachers an active role (something that a lot of MUS teachers like).

Two teachers highlighted the importance of reflection in a portfolio, connecting to the six participants who responded that reflection is important during the initial interviews. Nancy claimed, “Teachers should create a portfolio which can serve to show how the teacher has grown and reflected on his/her teaching. Time must be allotted for this process however, if it is not to interfere with the regular teaching/learning process in the classroom. This is a critical idea and should be built into the system in a meaningful way.” A strength of using portfolios to evaluate teachers is that it embraces a philosophy of contextualism, or giving meaning to observable facts (Ediger, 2000).

In the final group meeting, Loretta likened the portfolio to what MUS has been studying through its focus on visible thinking. By collecting evidence in portfolios and writing reflections, teachers made their thinking visible. It provided an insight into how they think about their teaching, which helped to document growth. Tracy also valued how sharing portfolios provided an opportunity to learn from one another, an example of Schein’s (1995) scanning. It provided an opportunity to see how her colleagues thought about teaching as well as what did or did not work for them. Nathan also agreed, but in
addition, valued how the portfolio process gave him a more detailed understanding of his students. He kept very specific data on his students’ mathematics progress and had them reflect as well. He said, “I found some satisfaction in knowing that information about my students in a very specific sense, which is not something I have been accustomed to doing.” He added,

I think kind of a culture at Mountainview Unified has been, with small class sizes, we perhaps have become overly confident that we know our kids. And we can always say that, you know, I know my kids. I know who the good writers are. I know who the intuitive math students are. And I felt pretty confident in that belief. But it kind of felt pretty good to be able to say, yeah, I know that and I also know some really specific details about them and about their progress or lack of progress.

The process helped him move from assumptions to facts. Tracy built on this, explaining that the portfolio made her feel accountable. She could not simply say she believed her students improved in their ability to identify the Habits of Mind in historical figures. She had to prove it with their work samples. Cathy also added that with the portfolio, she would have forgotten how she intended to change a lesson next time, or why a particular strategy did or did not work.

When asked in the final interviews if the process would have been as powerful without the portfolios, the group unanimously said it would not. They felt it was one of the critical components of an evaluation system. The only slight deviation from the resounding support for portfolios was a comment from Tracy, saying she felt she would have learned just as much about her own teaching without the portfolio, yet she would not have learned nearly as much from her peers if portfolios had not been utilized.
RQ4: After Participating in an Action Research Model to Develop a Teacher Evaluation System, how Satisfied are Teachers with the Co-Developed Teacher Evaluation Process?

Teachers were extremely satisfied with the co-developed model of teacher evaluation and believed it led to teacher growth and gains in student achievement.

As explained earlier, teachers felt that an evaluation system should help them grow and be motivating for them. Through reflections, group meetings, and individual interviews, I explored whether or not the system achieved these desires and teachers were satisfied with the process. Throughout the process, teachers expressed great satisfaction with the system, describing it as fun, joyful, invigorating, and more.

Teachers were more than satisfied with the evaluation system.

Though hints of teachers’ excitement about this new process were seen in earlier stages, clear support for the pilot system of teacher evaluation began to be seen during the mid-evaluation study group check-in meetings. Each participant expressed satisfaction and/or enthusiasm when asked how they felt about the evaluation pilot compared to a typical evaluation limited to formal observations. As Nancy explained, the process was,

Much better than having somebody come in and look because I would do the exact same thing as you [Tracy]. If you were coming in to observe me I’d plan a lesson that I’d done before that I know is successful and blah blah blah. And that doesn’t help me grow as a teacher.

In this one 30-minute meeting, Nancy used the word “fun” seven times when describing the process. “I’ve just had a blast!” she shared with her group. “I just love this process!” and later, “It’s just been so much fun.” Isabelle described how the process was motivating to her, saying, “It almost challenges me more than it did last year because now I know that I
need to, on a daily basis, continuously incorporate thinking routines or master my craft in various areas.” Loretta chimed in with, “It’s definitely been really valuable.” Cathy shared, “I really enjoyed it because it’s forced me to do things that I wanted to do. [LAUGH] And like kind of gotten me really focused and thinking of ideas and creating more.” Through these quotes and others, it was apparent that participants were satisfied with their involvement at this point of the study.

These same feelings were reiterated and expanded upon during the portfolio sharing meetings and final written reflections. It was clear that the participants enjoyed this process and felt it benefited both themselves and their students. All eight participants shared strong feelings supporting the value of the process we engaged in. Tracy’s response below is representative of the group’s feelings.

I just felt totally freed up to use it and not worry about if I was doing it right. That’s what I’ve worried about before. Am I gonna do it right? Do I really understand how this is supposed to be? .... It, it really, it’s exciting. I’d like to do some more. I said if I was coming back [not retiring], I’d be really excited about this process of working with people and looking at your growth. I think it’s amazing. It’s totally non-threatening.

Sarah valued the goal-setting process, saying, “That’s what makes this process so much more meaningful than the typical evaluative process. You get to set your own goals.” She elaborated in her written reflection with, “The process puts teachers in the drivers seat and gives us the ability to determine our area of focus for improvement which always boosts teacher buy-in.” Others saw this as a motivating step in the right direction, though they recognized there was much learning ahead of them. Ethan stated, “I feel like I’m just beginning this journey. I’m not using them [thinking routines] well in their journals or anything yet, but it’s kinda cool to have this foot in the water and get started.” Three other participants shared that this process forced them to do something they had been thinking
about, but just had not taken the time to try as of yet. Nancy was motivated because she felt successful. The process of collecting evidence showed her that her efforts made an impact on students. Her words describe these feelings of pride, saying, “It provides the process and structure for teachers to become more effective professionals, and in that, to grow to love teaching and to be proud of the important work we do for society.” According to Tracy, attitudes had shifted from counting down the days left in the school year to impatience for the next time to get together with peers and talk about teaching.

When comparing our developed model with a more traditional model of pre-announced observations, Ethan commented,

There’s no connection to that evaluation process, it’s sort of like, okay, my connection would be I need to find something that’s gonna really look good for that day. Versus this is something that’s actually [sic], I care about at this point. You know, and it was an idea that peers cared about.

This process connected the adults to the learning process, motivating them to engage in it and benefit from it. Sarah expanded on this thinking, sharing, “It made me connect to my practice.” By working with peers, bouncing ideas off of one another, and getting feedback, teachers were challenged to improve their practice.

Another advantage seen in this process was creating a safe environment to take risks and learn from peers. Tracy noted, “Knowing that other people are working with the thinking routines who can, you know, I can feel like I have back up before I jump off the diving board. It’s really helpful.” Nancy though, “Having a support group of teachers made it easier to take risks,” and Isabelle spoke of “safety in numbers.” Loretta said she was “inspired” when commenting on Isabelle’s presentation. Tracy also noted, “I was just thinking Loretta never shared that in the whole group. She’s always quiet. But here we are in a group of four and she’s giving us some new ideas, things we can use.”
It is clear from these responses and more that participants were satisfied with the designed system of evaluation. Cathy stated, “I think one thing I learned from this is that I wish I would have started it earlier!” Tracy, who is retiring at the end of the year summarized with, “If I were to be returning next year I would be very excited about this collaborative process. In my 46 years in education it is the most productive evaluative process in which I have participated.”

RQ5: What are the Challenges, and Lessons Learned in the Process of Teachers Co-Creating a Teacher Evaluation System?

Many lessons were learned along the way that can help others successfully replicate a similar process in their own schools and districts.

Throughout the process of co-creating a teacher evaluation system, lessons were learned and challenges were faced along the way. By reporting these below, others can learn from the successes we saw and obstacles we faced. Participants made specific recommendations for the future of teacher evaluation and reflected on components of a system that were critical to its success. Furthermore, my reflection on the system and process illuminated many key ideas that should be considered when implementing systems of teacher evaluation.

Participants identified specific logistical components that they felt would lead to success as well as key elements of a successful evaluation system.

During the final group meeting participants were asked to clearly define a system of evaluation that they would recommend to others. The details agreed to were then verified individually in final interviews. Though they initially expressed reluctance to put hard numbers to the flexible system they so valued, they agreed to the recommendations below.
An annual twelve-week timeframe with four meetings would support teacher growth.

Participants came to consensus that an evaluation cycle should last approximately twelve weeks. Tracy shared that the longer the cycle became, the more likely it would create a competitive culture. With significant time to assemble portfolios, teachers may feel pressure to create more sophisticated and expansive presentations. The brevity of a twelve-week protocol would relieve any undue pressure to create works of art and compete with one another. However, Nancy shared in her interview that if it were too short, there would not be much to reflect on.

The group felt that the process should start with a goal-setting meeting where participants develop their professional growth plans collaboratively. They would then have two mid-evaluation check-ins (one after four weeks and one after eight weeks) to share progress and brainstorm ideas. The final meeting would take place at the end of the twelve weeks with study group members sharing their portfolios with one another. Isabelle explained that in addition to the four meetings, there would probably also be additional meetings that happen organically as particular participants have needs they would like to discuss. Staff members also felt that three to four members was an ideal size for study groups.

When asked if this process should occur on the typical California cycle of evaluations occurring every other year, Isabelle responded, “I think teachers should learn and grow every year. I don’t think you can argue that.” Heads nodded around the room and the group agreed that if the administrator had the time to manage being involved in
this process with each employee every year, they would enjoy doing this on an annual basis.

**Participants felt they system should include goal setting, peer observation, and evidence collection.**

As I have articulated throughout this chapter, participants clearly valued goal setting, peer observation, and evidence collection. In addition to the observations throughout the research, participants highlighted these three key traits in their final interviews when asked what components an ideal evaluation system should include. Seven of the eight participants mentioned goal setting when asked this question. All eight mentioned peer collaboration and seven mentioned evidence collection as critical components to the success of an evaluation system.

**Precision of language and clarity are highly important.**

With earlier sentiments on the importance of feeling safe, it was important for participants to understand the purpose of the evaluation and the administrator’s motives. As Nathan explained, “Administrators or whoever the other evaluators are, ... need to take special care in-in terms of making sure teachers understand their motives.”

Though noted in the initial interviews, the power of language and pitfalls associated with misinterpretation were obvious in the initial group meeting. In particular, the teachers did not feel good about the word “evaluation.” Nancy shared:

In a way, you almost want to take the word evaluation out of it. If you take the word evaluation out of it, I don't think people will feel threatened, you know, because that word just has a connotation to it. So we can come up with some other word for that because I agree with what you’re saying. To me it’s all about growing as a teacher, you know, and really looking at your own practice and celebrating what you’ve done and how you've come from the last time you did it.
Later, when attempting to avoid the word evaluation, I used the phrase “do better,” and the following conversation ensued:

Tracy: I like to hear you say to grow instead of do better because when you say do better than that, it implies they’re not doing well.

NB: So the language is important.

Nathan: Yeah.

Nancy: It is important.

Tracy: It’s very important, I think.

In addition to the choice of words, I also discovered that it was critical to ensure that the group had the same understandings of the terms we were using. When the idea of portfolios was first presented, it was met with resistance because some group members were envisioning the portfolios that they created when completing teacher education programs or applying for jobs. Nancy said, “I see that [when you say] portfolio, some people go like, AHHH!” The portfolios they were envisioning were large collections of lesson plans, student work, pictures, and teaching philosophy statements. My understanding was much more loose, and I clarified, saying:

I think to me there is some confusion around the language of portfolio. I guess when I say portfolio it just means any collection of evidence. Not necessarily a three ring binder with specific pieces stuffed in. So, you know, videos, self-reflections, student work, any piece of evidence that would say, “this is how I got better at teaching math.” Or “this is a thinking routine I tried.” Anything could be used as evidence to say, “This is how I grew towards these goals.”

This clarification caused a definite shift in attitude towards portfolios. Cathy expressed her feelings, stating, “once we have a common understanding of what portfolio means then it’s not so scary.” Sarah agreed. “The portfolio seemed like, oh, that is such extra work and hoops to jump through, but the way that you’ve explained it, it doesn’t seem that way, so with explanation maybe you’d have more buy in.” A similar experience
occurred in our discussion of the use of value-added measures in evaluation. When I asked if any of the means of evaluation did not align with teachers’ beliefs at MUS, the only item mentioned was the use of value-added data. Some of the teachers understood value-added to mean judgment based on end-of-year state test scores. After clarifying that value-added could mean any growth in student success demonstrated through teacher-collected data, teachers were more comfortable with the use of value-added data as a piece of evidence in a portfolio. The power of precise language and clear understandings was visible in these interactions.

**The process of reading the literature review changed and/or strengthened teachers’ beliefs about evaluation.**

Participants reported changes and affirmations in their beliefs about teacher evaluation after reading the literature review. Cathy spoke to her perspective opening up on what could constitute an evaluation. Loretta in particular noted how reading the research and answering journal prompts caused her to reflect on her experiences and her desires.

After reading the literature review, which caused me to constantly consider what I had experienced against what I was reading, I feel as though it definitely opened up new lines of thinking. It was invigorating to try to conceptualize what a better model might look like, because to think of the end result - fine-tuning & bettering myself as a teacher - is exciting. I feel as though my views on best practices in teacher evaluation have opened up, and I am able to consider a broader scope of the purpose as well as a variety of means in which to pursue it.

Other teachers spoke to how the activity strengthened beliefs they already had, now knowing that they are based in research. Sarah noted

I think if anything the views I already held are just solidified. I believe even more strongly that student achievement can be a useful tool but must be used with open
eyes and extreme caution. I am still convinced that MORE observation is better and agree completely that peer observation and review should be used regularly.

Nancy had a similar experience. However, she experienced some change as well, noting, “I am now aware of the need for numerous components of the evaluation process; clearly one or two formal evaluations by an administrator do not paint a complete picture of the teacher being evaluated.”

**Participants were more likely to change their goals if they did not come with the forms already filled out**

In preparation for the study group pre-conferences, I asked participants to begin to think about a possible goal and forms of evidence they might use to demonstrate progress towards that goal. Three of the participants arrived with professional growth plans completed and typed (Nathan, Loretta, and Nancy). Two had handwritten notes (Isabelle and Cathy) and three more arrived with nothing written (Tracy, Ethan, and Sarah). Those that had not formally typed up their professional growth plans demonstrated more openness to altering their plans based on feedback from peers and myself. There was no change between the initially presented goals and the final agreed upon goals for the three participants who came with pre-typed plans. Isabelle, who had handwritten notes, also demonstrated no significant change from her original orally presented goal. However, each of the other participants demonstrated various changes between their originally presented goals and what was finally agreed upon.

The types of changes demonstrated in goals varied from clarifications of language and desired outcomes to outright course-correction and exploring goals that had not been considered prior to the meeting. Tracy was an example of a participant who clarified and refined her goal through peer interaction throughout the meeting. Before even stating her
idea for a goal, she began by sharing her desire for help in thinking about her goal and refining it. Her initial goal was to use thinking routines throughout Black History Month and study the Habits of Mind of historical figures. Tracy's motivations and thinking were clarified by others asking, "Why do you think it's important? Why is it a worthwhile endeavor to know about the Habits of Mind?" Through these questions, Tracy was able to clarify her desire to build students' awareness of the Habits of Mind and teach them to identify the habits in others as well as themselves. She also expressed that her desire to use thinking routines was based on wanting to build her own comfort and fluency with the routines so that they could be more fluidly and spontaneously incorporated into her lessons. From her vague initial goal, it evolved to:

Develop student’s understanding of Habits of Mind and ability to identify them in historical figures and themselves through a study of Black History Month while also building my own fluency with thinking routines.

The others who demonstrated change also had similar shifts, but Ethan changed topics altogether during the meeting. He came in assuming he would talk through a goal focused on language arts, but after hearing two colleagues speak about working on thinking routines in mathematics, he changed his focus.

It's sort of funny because, sitting here, listening to you I'm thinking, you know, it [LAUGH] would be kind of neat to also do, because for me, I have done almost no thinking routines in my math, and [LAUGH] I'm just sitting here thinking, I'm being intrigued with the idea of doing math algorithms. [LAUGH] I don't have as many layers- probably that I want to do as you do ... but it would really be intriguing to me to maybe even do that at the same time, especially if you are thinking about doing it, that would maybe make sort of an interesting discussion for us.

---

8 From Harvard’s Project Zero Classroom, classroom routines designed to cause students to think in new, novel, and critical ways.
9 From Costa and Kallick’s Institute for the Habits of Mind focusing on 16 essential skills, or habits, that help people be successful.
Ethan was presented with disconfirming data. He had not reflected on his lack of thinking routines in mathematics until hearing that others were planning to use them. He also hinted that he was drawn to the psychological safety of exploring this area together with peers.

Although Isabelle was the only participant who had not typed her professional growth plan and did not significantly alter her original goals, it is important to note that she was a close colleague of Loretta, who had typed her plan. Isabelle and Loretta work in the same grade level and essentially proposed the same goal, using more thinking routines in mathematics. The two had engaged in discussions about their goals prior to the preconference, which might explain why her goal did not change.

**Participants felt the action research process led to increased buy-in to the co-developed teacher evaluation system.**

Of the eight participants in this study, each felt the system would not have been as effective if the group had not participated in the action research process together. The process of reading the literature review and co-creating the pilot system led to a deeper understanding of why the system should work, serving to motivate teachers to engage in the process thoughtfully. Sarah shared,

> When I was reading those things [in the literature review] I was thinking, oh, yeah, this [the old system of evaluation] is so unjust. This is such a terrible way to evaluate teachers. And then I started thinking about how I've been evaluated in the past, and it hasn’t been negative, but it hasn’t really been helpful. And so really thinking about the evaluative process before we launched into it kind of got me already thinking this isn’t a good way that it’s been done, and well, I’m excited to see how we could do it differently.

Nathan expanded on Sarah’s thinking, saying, “The opportunity to spend some time reading, and perhaps do a little of your own research and kind of form those foundations of
knowledge before going forward, I think, were really essential.” Nancy shared that being
presented with the literature review was much preferable to having to do all of that
research independently. Isabelle articulated her thoughts, saying,

I’m not so sure how much buy-in it would’ve gotten had you gotten the eight of us
together and then said, “Here are some final products; here’s what we did last year.
What would you like to do this year?” I think it might’ve looked a little more – I
don’t know, intense is the word that comes to mind – but because we were involved
from day one in the building of it and what it might look like, and there was so much
flexibility in-between, then that felt really comfortable. And then the staff just took
it from there and excelled.

Tracy compared the process to including students in the creation of class rules. “It’s like
the class. If they write their own rules they’ll be more likely to follow them.” They did not
believe that our final product could be picked up and dropped off at another school and
find the same success. Ethan explained, “You build that trust by having that common
understanding of people and how people grow and learn. And then when you know that
that’s there, then you can trust that, okay, I can take that next step.” Being involved from
the initial stages led to motivation and buy-in and developed the psychological safety
needed to take risks and try something different. They were invested in the success of the
system.

The staff members were initially concerned about the time and resources that
would be necessary to support a system of teacher evaluation, yet did not feel this
process was a burden when reflecting afterwards.

The group clearly agreed that an evaluation system should not be an unreasonable
burden on their time and energy. In the initial planning meeting, they were anxious to take
advantage of systems already in place at the school, such as professional learning
communities, to accomplish this work. In professional learning communities, grade levels
and special interest groups meet regularly to explore areas of interest. Because of our flexibility in evidence, the group quickly saw how any work with PLCs that related to one’s evaluation goals could serve as part of the portfolio. This was important to Nancy, because Tracy did not want the portfolio process to draw away from energies needed for the classroom. However, I explained that if teachers were truly learning and improving from this process, the energies would not be taking away from the classroom, but in fact might provide for even greater efficiencies. If teachers truly saw this as a valuable experience, there should be no worry about the extra energy expended because it would have immediate impacts on their effectiveness in the classroom.

The concerns about workload were not present in the final group reflection and individual interviews. When specifically asked about the scope of the work required for this process, no teacher felt that the process was a burden or undue hardship. Cathy explained how the process itself was motivating enough so that it did not feel like a burden at all. She shared,

It would be really hard to not get motivated once you started the process. I think you’re going to choose something you want or hopefully there’s something that interests you, if there’s not, then that’s a problem. And then, once you start going with it, if you’re not into it, then maybe you shouldn’t have chosen that, or there’s another problem … I feel like people are going to take off with it.

Because the process was engaging and motivating, it did not feel like extra work to Cathy.

Efforts towards achieving evaluation goals infiltrated other aspects of teachers’ work.

Although participants’ goals were clear and focused, four of the seven participants explained how their experiences in this evaluation process started to permeate other areas of their teaching at the mid-evaluation study group check-in meetings. The learning was
not limited to the scope of their professional growth plans. Nancy described how her work with thinking routines and Black History Month led to exploring thinking routines in mathematics, saying “but, what was really fun [was] that math, the math stuff that we’ve done because I didn’t expect to go into math, you know . . . it’s like – it’s exploded into other areas.” Loretta explained how her current work in the pilot has “set the stage” for students in other upcoming units. Not only did her work on this particular math unit infiltrate other units, but it sparked an awareness of thinking about mathematics. She explained that she now “always has her ears open” to new ways of thinking about mathematics. Cathy shared that the process has actually acted as a catalyst, inspiring her to share more of her work with her grade-level teammates because she now thinks about her work more deeply and wants to hear their perspectives on what she is doing as well as share her successes so that they might replicate them. Tracy’s experience seemed to constantly inspire her to revise her teaching. “I just started with a simple plan and it just kept evolving.” She spent considerable time revisiting how one lesson, discussion with a peer, or time spent reflecting led to new ideas. Her excitement was captured when she said, “oh! and then I had them do this about each other [describe one another with the Habits of Mind]. . .I mean, that’s what I said, it just keeps going on and on.” It was clear that for some participants, the effects of this process went beyond the mere scope of their original goals.
CHAPTER 5

The purpose of this study was to determine what teachers at MUS believe about teacher evaluation as well as identify methods that could be used in an evaluation system that motivate them to learn and grow. The study identified steps, potential pitfalls and lessons learned when engaging a group of teachers in action research to design a system of evaluation.

This chapter begins with a summary of findings, then walks through a discussion of a visual model describing conditions that lead to teacher growth and gains in student achievement. After revisiting connections between the discussion and research questions, conclusions are drawn and recommendations are made for policy and practice. This chapter ends with my plans to disseminate the findings of this work.

Summary of the Findings

This study provided insight into what teachers believe about teacher evaluation, and what systems and structures best support teacher growth. Given the right conditions, teachers will be motivated to engage in their practice, take risks, reflect, and refine how they teach in an ongoing manner.

- Teachers believe that the purpose of teacher evaluation is to help them grow as educators. This can be accomplished by providing a safe and supportive culture and systems that motivate and engage while remaining flexible enough to meet the needs of diverse teachers.
- Teacher growth is supported through various methods, including peer interaction, goal setting, evidence collection, reflection and an ongoing and iterative process.
• Teachers reported being highly satisfied with the co-developed model of teacher evaluation and believed it led to teacher growth and gains in student achievement.

• Many lessons were learned along the way that can help others successfully replicate a similar process in their own schools and districts. Most importantly, specific logistical parameters were recommended, the importance of precise language and clarity was seen, and the process of reading the literature review changed teachers’ beliefs about evaluation.

These major findings have significant implications for the future of teacher evaluation. With the current national debate on value-added assessment data and federal funds being designated to states and districts that adopt certain methods of evaluation, there is no better time to fully understand what motivates teachers to grow so that we can better impact student achievement.
Discussion

Figure 5.1: Model of Teacher Growth

I developed the visual model above to help explain why I believe the system and process we used led to teacher growth and satisfaction. Generally, this model of evaluation puts teacher growth at the center and assumes student achievement will follow when checks and balances are in place. However, to get teacher growth, Schein's (1995) five components of adult learning must be present. Psychological safety is of utmost importance and can be created through structures and activities. Peer observation, goal setting, evidence collection, and frequency all serve as motivational forces that provide opportunities for disconfirming data, scanning and cognitive redefinition as well as psychological safety. To simplify, I will walk through various subsections of the model and speak to each facet individually.
Teacher Growth and Student Achievement

With the current national debate, the issue of value-added measures of student achievement must be addressed. The LA Times recently published standardized test scores for individual public school teachers and the New York State Court of Appeals ruled that New York teachers’ performance assessment results could be made public. The national Race To The Top program has now opened up its federal fund grant applications to school districts, not just states, so long as a part of teacher evaluations incorporates student test data. Politicians, union leaders, and even Bill Gates are making public statements regarding the use of student test data to evaluate teachers. In an Op. Ed. piece to the New York Times, Gates (2012) says that student test scores alone are not a sensitive enough measure of effective teaching. He stated, “A good personnel system encourages employees and
managers to work together to set clear, achievable goals.” He goes on to say, “Developing a systematic way to help teachers get better is the most powerful idea in education today,” [my emphasis]. This is exactly what we did in this study. We developed a systematic way to help teachers get better.

The realm of public education is often a polarized environment where administrators and teacher unions lack trust in one another. The evaluation of teachers therefore becomes a point of contention where teachers feel the need to protect themselves from capricious and arbitrary evaluations. Administrators feel crippled by the current systems of evaluation, which do not give them a deep understanding of a teacher's abilities, nor does it give them the tools to help the teacher learn and grow. Because of this, it is easy to look to value-added measures of teacher effectiveness when one does not have any other measures at his or her disposal. However, the research presented in the literature clearly states that value-added measures are unreliable and frequently wrong. Many do not believe that the tests measure what is actually important for students to learn at school. Furthermore, a large number of teachers have students who do not take standardized tests because of their particular grade level or subject. Participants in this study unanimously agreed that current standardized assessments are too fraught with imperfection to be considered in teacher evaluation. Even if assessments were improved to included better measures of critical thinking and problem solving, participants still felt they should contribute to a minor percentage of an evaluation as a whole, with all but one response falling between 0% and 33%. Instead, teachers felt a more substantive and qualitative process such as the using portfolios would be more effective.
Instead of trying to fix a broken system by including value-added measures, we need to replace the entire system of evaluation. Moving towards a teacher-centered model of evaluation is a critical first step in repairing the relationship between administrators and teachers and creating cultures of trust, safety, and risk-taking. Teacher evaluations are currently a lost opportunity. The stakes are too high to let this opportunity continue to go to waste.

The participants in this study felt the key to teacher growth, and thus to increases in student achievement, is creating a supportive culture that motivates teachers to think about their practice, seek new learning and understandings, build relationships with colleagues, experiment, and enjoy their work. They believe this model accomplished each of these. They felt strongly that the use of value-added measures, teacher rankings, and teacher grades would erode and undermine the culture needed for teachers to be in a state ready to learn and grow.

How can we know this system promotes student achievement if we do not include standardized measures of student achievement? When contemplating whether or not this structure truly makes a difference for students, we need to examine how those conclusions are drawn. The process we used had a built in system of checks and balances. Observations, written reflections, local assessment data and student work were all used to judge whether or not students (and teachers) made growth. This perspective also included a shift in the judgment of adequacy coming from administration to the onus being placed on the teachers. The study group took the stance “convince us” and asked the individual teachers to demonstrate why they believed students made progress. This shift in control is important as it motivates the teacher and empowers them to make a difference. They do
not have to rely on an external evaluator to tell them whether or not they are doing a good job. They know it intrinsically based on their reflection and analysis of student work. They feel proud. They feel capable. They feel inspired to continue down this path of learning and growth. Rather than dreading the bi-annual evaluation, teachers are asking for more and hoping the process can expand.

Skeptics may think the thoughts above sound good in theory, but how would we know it is truly making a difference for students? Analysis of participants’ experiences leaves no question that this work had an impact on students. The system incorporated checks and balances, including student work samples, reflections, observations, and local assessment data. These checks and balances can answer the question above, providing evidence of real changes in students’ abilities. As shared in the findings, the stories of growth presented at the portfolio sharing meetings clearly describe powerful student achievement that transcends improvement on standardized tests. Cathy told the story of John who did not see himself as a learner and would give up easily. By sharing writing samples as evidence of his newfound can-do attitude and video of him talking about his persistence, it was obvious that Cathy had changed the future of this child’s education. Not only did she impact this one particular student, she also walked the study group through example after example of each student’s growth in writing in six short weeks. These kindergarten students as well as first grade students do not take standardized tests and there is currently no means of calculating a value-added score for these teachers. Rather than potentially missing opportunities to improve the first two years of a child’s education by using the antiquated system of evaluation, this evaluation system led to a specific
change in a teacher’s practice that will impact the educational journey of these young students.

Despite the generally flexible nature of the system, there were mechanisms in place to ensure the work was focused and appropriate. By running ideas through study groups consisting of peers and administrators, a system of checks and balance was incorporated. If a teacher were to come with a goal to teach students to improve their juggling skills, the administrator and peers would have the opportunity to question the appropriateness of the goal. In regards to goals and evidence, the position of “convince me” was utilized when entering a conversation. Perhaps a physical education teacher could convince the group that teaching students to juggle connected to standards such as hand-eye coordination and gross motor skills and would be deemed appropriate.

In addition to peer review as a means of ensuring the flexible nature of the system was not taken advantage of, two pieces of evidence were also required of all participants. Each participant was required to have an observation component as well as reflection. The observations could be scheduled or unannounced and could happen with administration and peers. This ensured that someone else saw parts of the teaching process. By having each participant reflect, it also provided the group with insight into how the teacher thinks about teaching and what value they were finding in the process.

The use of a value added measures may shame teachers into wanting to improve and provide administrators with “evidence” to validate high-stakes employment decisions, but this is not how we will change the face of education. According to Tracy, attitudes at MUS shifted from counting down the days left in the school year to an impatience for the
next opportunity to collaborate with peers and talk about teaching. This is clearly what we want an evaluation system to accomplish.

This model operates on the premise that teacher growth will lead to student achievement. One is the natural cause of the other. When teachers do a better job teaching, students will benefit. The best way to improve one’s teaching is to study it, to actively think about it and talk about it with peers. Like gravity, gains in student achievement are inevitable when teachers study their practice and work to improve their teaching.

**Figure 5.3: Psychological Safety**

![Psychological Safety Diagram]

**Creating Psychological Safety**

The most critical component to a system that promotes teacher growth is creating psychological safety. For adults to learn and grow, they must feel safe to take risks, to change, and to question their ways of thinking. Numerous actions can take place and structures can be implemented in order to create this psychological safety.
An evaluation system aligned with teachers’ conception of teaching can increase psychological safety.

Teachers can feel safe when the method of evaluation aligns with what they believe about evaluation. Teachers in this study clearly stated that they believe teaching is a profession. In Darling-Hammond, Wise and Pease’ (1983) framework, teachers who view teaching as a profession value using their judgment to respond to the unique needs of students and situations. The pilot system we developed was designed to allow for, and honor, their judgment. They had flexibility in choosing their goals and methods of collecting evidence. Also, by including reflection as a critical piece of evidence, teachers were forced to articulate how they used their judgment to approach their goals in the way that they did.

Isabelle shared that the biggest difference between our pilot system and the traditional system is that she felt treated like a professional. Ethan saw this as a critical component as well, saying,

You’ve got to give something to them to say that it’s not a job that you’re just here doing, but it’s a profession in which you can be growing and thinking and your intellect is connected to what you’re doing.

In his final interview, Nathan shared, “I view professionalism in both the academic context and also in a larger context, kind of being proud of what you do, and being willing to dedicate yourself to it.” The process we used allowed teachers to feel proud of their work and gave them a forum to share their growth, honoring their professionalism and working to create safety in the alignment between evaluation and teaching beliefs.
Psychological safety can be enhanced when the administrator acts as a facilitator.

Participants agreed that my administrative role as a facilitator helped create a climate of safety and trust that supported and encouraged growth in a non-threatening way. By remaining neutral throughout the process and not inserting personal judgments, administrators can work to promote psychological safety. Ethan emphasized the importance of creating an environment where people feel empowered and supported. His ideal role for an administrator would be, “helping people think about their process and helping them think about what they’re doing, and where they’re going.” Though it can be difficult to separate one’s self from the traditional hierarchical role of an administrator, the way in which an administrator conducts him or herself can communicate clear intentions to teachers. Through actions and words, teachers in this study demonstrated that they could be convinced that I was there as a true support, not someone who holds the knowledge or renders judgment. They felt supported through my role as a facilitator.

By flattening the hierarchy between administrator and teacher, trust was gained. Participants were involved in the decision-making process. They were empowered, and entrusted, to design the new system. In my initial interviews, a number of participants explained that during prior evaluations they simply did what they thought would make the administrator happy. They put on a show to fulfill their supervisors’ expectations. The administrator held the power and controlled the system of evaluation, often making the process feel like jumping through hoops. Involving teachers in the process empowered them. This control and influence helped increase psychological safety. It signaled that their ideas and beliefs were valued.
Co-creating a system of evaluation through an action research process creates increased psychological safety.

The act of reading the literature review and co-creating a system of evaluation promotes clear understanding of the evaluation system, the intentions of it, and the value of it’s structures and processes which leads to psychological safety. Participants in this study expressed how critical they thought this process was to getting buy-in from them and they recommended that all schools go through the process together. Lack of knowledge can breed fear, so knowing why an evaluation system is designed the way it is can increase psychological safety. Furthermore, by co-creating the system with an administrator, there is further opportunity for the teachers to be convinced that the administrator’s beliefs are aligned with theirs. The transparency of the process leaves no room for doubt, helping the participants to feel safe taking risks and fully engaging in the process.

Demystifying teacher evaluation by creating common understandings.

Another benefit of the co-creation of an evaluation system is that it creates common understandings and agreements. Through this process, evaluations were demystified. Not only were teachers unsure of how an evaluation could and should happen, but they also were not clear on what the purpose of evaluation should be. How can we expect a tool to be valuable with so much uncertainty? Furthermore, how can we expect teachers to feel safe enough to take risks and push their thinking when they have little to no control over the process? Many aspects of this research worked to bring clarity to these issues and create common understandings among myself and teachers, building a climate of trust and safety that promoted risk-taking, reflection, and growth.

Understanding existing beliefs.
In order to create a system of evaluation that teachers understood and believed in, we needed to understand what we believed about teaching and the purpose of evaluation. We also needed to do this in a manner that would work to create a climate of trust. This began by hearing teachers’ beliefs individually in pre-interviews, then coming to common understandings in group meetings. It quickly became clear that teachers thought of themselves as professionals, though they sometimes operated in Darling-Hammond, Wise, and Pease’s (1983) *craft* and *art*. Furthermore, even though they saw themselves as professionals, they were attracted to being artists occasionally, though teaching as a craft often had negative connotations and was seen as beneath teachers at MUS. Because they primarily saw themselves as professionals exercising judgment and valued occasional artistry where intuition and flexibility is key, it was obvious that a standard checklist evaluation would not meet their needs. For them to learn and grow, they needed to have professional conversations about their practice. This knowledge was critical in helping design a system they would be satisfied with. Additionally, providing teachers with Darling-Hammond, Wise and Pease’s framework gave them a new paradigm with which to understand their work and engage in a productive conversation about methods of evaluating teachers.

*Co-creating new understandings.*

Although understanding their existing beliefs was important, the process of co-creating new understandings of ways to evaluate teachers was also a critical step in this research. This was done by having the participants read a literature review on teacher evaluation. This process provided what Schein (1995) refers to as *scanning* and *cognitive redefinition*. Scanning involves being exposed to new ways of thinking by reading articles,
taking classes, or observing others. Cognitive redefinition refers to understanding new ways of thinking by having broadened interpretations of words and concepts, revised anchors or scales of judging, and altered emphases on what is or is not important. Grounding the scanning process in research also added credibility to these new understandings that helped motivate participants to accept new ideas. If I had simply asked participants to collect portfolios containing evidence of progress towards their goals, they may have balked at the additional workload and been predisposed to doing the work with disinterest and disdain. However, after reading the literature review and research on the value of portfolios, participants were advocating for them and conscious of the benefits as they assembled and presented them, giving more meaning to the act. They experienced cognitive redefinition. This redefinition was further solidified when participants had the opportunity to discuss their thoughts with one another in the planning stage. When someone hears a number of his/her peers commenting on the value they see in using portfolios, they are more likely to align themselves with the others. Schein refers to this as disconfirming data, information that causes one to reevaluate one's beliefs. This cycle of being presented with disconfirming data, scanning for new information, and cognitive redefinition is a powerful force in changing behaviors. Be that as it may, change is hampered if people do not feel psychological safety. Without feeling safe enough to take risks and change, disconfirming data is often ignored or external blame is used to invalidate the data.

*Clarifying the purpose of evaluation.*

Clarifying the purpose of evaluation helped create the psychological safety needed for cognitive redefinition. In the past, teachers did not know why they were being
evaluated. Some thought it was to weed out weak teachers, some felt it was just an administrative box to be checked, and others thought it was an opportunity for administrators to tell teachers what they need to fix in their teaching. By hearing their thoughts and coming to explicit understandings of the purpose of this system, they were relieved from the pressure to perform perfectly during a typical observation. By defining the purpose of evaluation as helping teachers grow, it gave them freedom to take risks, experiment with new techniques, and ask questions. This safety promoted vulnerability and openness to new ideas, without which change would not occur.

The flexible nature of the developed system led to psychological safety and satisfaction.

When agreeing to the use of portfolios in this study, teachers appreciated the flexible nature of goal setting and the evidence collection process. If a teacher was not comfortable being videotaped, they did not have to choose it as one of their pieces of evidence. Teachers who preferred to work alone could document their lesson development process and independent research, and reflect individually afterwards. Teachers who thrive on interaction could document conversations with peers, experiment with collaborative teaching models, and reflect on the strengths and weaknesses of their efforts with a critical friend. The goal setting process mirrored this, with three teachers adopting basically the same goal on thinking routines in mathematics while other individuals primarily worked in isolation. By providing choice, participants felt safe and supported. Again, the structure supported Fairtlough’s (2007) notion of a responsible autonomy where individuals have the freedom to make choices, yet are also accountable for those choices. It also aligned with Theory Y beliefs that employees enjoy work and can creatively
problem solve when trust and cooperation exist between administrators and teachers.

Finally, the flexible nature of the system paralleled key traits of teaching as a profession and teaching as art. Professionals and artists exercise judgment and intuition compared to the more inflexible craft and labor view of teaching, all working towards increasing psychological safety for participants.

**Figure 5.4: Motivational Forces Leading to Components Required for Changes in Adults Behaviors.**

![Diagram of Motivational Forces](image)

**The Four Motivational Forces**

Peer interaction, goal setting, evidence collection and an ongoing and iterative process provide opportunities for adults to experience psychological safety, receive disconfirming data, feel survival anxiety, scan for new information, and cognitively redefine
their beliefs about teaching. These forces all interact to create the conditions necessary for adults to change their behaviors, leading to teacher growth.

**The power of peer collaboration.**

In addition to the benefits of peer collaboration found above, peer collaboration was also a critical piece of the newly designed system itself.

**Peer collaboration is aligned with teaching as a profession and art.**

When thinking of teaching as labor or craft, an external evaluator would be enough to verify if teachers are following scripts appropriately and implementing basic strategies in teaching. As Weems (2010) noted, if the purpose of an evaluation is to determine competence, an observation by an administrator is enough to accomplish this. However, when considering teaching as a profession or art, teachers are valued for using their judgment and intuition. A simple observation does not provide adequate insight into why decisions are being made as well as the thought that goes into designing a lesson and unit and the reflection that occurs afterward. Furthermore, when teachers’ judgment and intuition are valued, it makes sense to leverage these abilities by incorporating peer interaction throughout evaluation and not rely solely on administrators to dictate the process, controlling the knowledge of what is or is not good teaching. This was seen throughout the pilot, with teachers seeking advice from their colleagues and incorporating the ideas of others into their own plans.

**Peer collaboration creates circumstances that promote learning and growth.**

Schein (1995) believes that for adults to change behaviors, they must feel psychological safety, be presented with disconfirming data that brings forth survival anxiety, and have opportunities to scan for new information in order to experience
cognitive redefinition. Embedding peer collaboration into a system of teacher evaluation creates opportunities for each of these to occur.

Peer collaboration promotes psychological safety.

When adults are observed by their peers, it lowers the anxiety relative to being evaluated and observed by an authority figure (Britton & Anderson, 2009; Iwanicki, 2001; Kohut, et al., 2007; Marshall, 2005; Wilkins, et al., 2009). In this model, participants were “observed” by peers not only during their teaching, but also during their interactions in goal setting, mid-evaluation check-ins, and evidence sharing. Of the eight participants, seven mentioned the need to feel safe and supported as critical to an evaluation system and six participants spontaneously offered peer observation as a valuable tool to be used in teacher evaluation during the pre-interviews. The less hierarchical relationship amongst a group of teachers and an administrator is much less threatening than an exchange that takes place between a single teacher and administrator. Without this trust and psychological safety, the positive outcomes associated with peer review may not be realized (Britton & Anderson, 2009; Byrne, et al., 2010; Iwanicki, 2001; Kohut, et al., 2007).

Peer collaboration provides opportunities for disconfirming data, which in turn induces survival anxiety when psychological safety is present.

When peers interact, they are exposed to new information. If this new information makes them uncomfortable enough to challenge them to rethink their assumptions, the disconfirming data leads to survival anxiety and an impetus to change. Survival anxiety is the feeling that one must change in order to “survive” (Schein, 1995). In the teaching realm, survival could mean maintaining a reputation among one’s fears or preserving one’s self-esteem. For example, observing a colleague who exhibits stronger classroom
management than one’s self could provide disconfirming data and a sense of survival anxiety. The teacher may feel that they are not a “good” teacher when compared to this peer and be motivated to improve. Observation alone is not the only way to receive disconfirming data. The dialogue that took place throughout this study might also serve as disconfirming data. Teachers could be caused to feel survival anxiety if they noted that others are thinking about teaching to a level of complexity and depth that they are not. Outputs such as professional growth plans and evidence also provided disconfirming data in this study. When Sarah arrived at the initial professional growth plan meeting she was the only participant in her group that did not have her thoughts detailed on paper. She made a quip about being in the gifted group and it was obvious that she was uncomfortable being an outlier in the group. She felt survival anxiety and wanted to save face with her peers.

*Peer collaboration provides opportunities for scanning, leading to cognitive redefinition.*

Once someone feels survival anxiety, they are primed to scan for new information in order to avoid failure. Scanning can occur by observing peers, engaging in dialogue with perceived experts, or other means of research. When Nathan was dissatisfied with his students’ growth in addition and subtraction fluency, he went online to better understand how mathematical facts get into long-term memory. When Sarah understood that parent involvement was critical to her students’ success in violin, she asked her study group for help. These are explicit examples of scanning, but it also occurs much more subtly as well. A teacher who is reluctant to be videotaped sees her peers finding value in the process and decides it is worth a try. She scans cognitively redefines her notion of what she is
comfortable with based on the beliefs and practices of others. Ethan noted that two of his colleagues were focusing on thinking routines in mathematics and decided to join them.

Learning also occurs through imitation. Ethan likened this process of cognitive redefinition to students copying one another in class.

The first six weeks you're going to find that a lot of kids are just looking at what the kid next to him is doing, and doing the same thing, but after five, six weeks; all of a sudden they're starting to do their own thing. And, I think that would happen with folks that are struggling, if they're in a group.

When peers engage in meaningful conversations and share their interpretations of lessons and teaching activities with one another, they are scanning for new ideas and coming to new understandings.

**Peer collaboration assumes a Theory Y model of motivation.**

A failure of the traditional model of teacher evaluation is that it relies on the validity of MacGregor’s Theory X (Northouse, 2007). With Theory X, administration assumes teachers are inherently lazy and dislike work. It is the administration's job to monitor them. Under Theory Y, administrators believe that teachers should enjoy their work and they can be entrusted to creatively problem solve. Peer collaboration empowers teachers to work together to solve their problems, not relying on an authority figure to tell them whether or not they are doing a good job.

**Goal setting and evidence collection.**

When initially asked about the purpose of evaluation, a number of teachers in this study felt an evaluation system should motivate them, inspiring them to learn and grow. The process of setting goals and collecting evidence may have contributed to motivating
teachers to change. Nancy shared how a goal setting and evidence collection process was motivating for her,

Rather than someone else telling you, you have to do this, then the teacher would figure out what they had to do. It gives you power. And it, and makes you feel proud at, when you come in and you have it all, you’re kind of giddy. It’s pretty fun!

Earlier in the study, three other participants agreed that involvement and ownership in the process through goal setting would be motivating to them. These beliefs were reaffirmed by participants in the final interviews, where seven of the participants spoke to the power of goal setting and evidence collection as critical to the success of the pilot.

**Goal setting and evidence collection create psychological safety.**

Within a tradition evaluation system, teachers are often left guessing what the administrator would want to see in an observation as well as anxious about the observed lesson not going perfectly. By setting an explicit goal or goals, teachers know exactly what to focus on and demonstrate, creating psychological safety. Furthermore, the act of collecting evidence allows multiple opportunities to demonstrate proficiency and growth towards the goal. The stakes are lowered because a teacher does not have to prove him or herself in a single observation. In fact, the administrator does not even have to be present to witness progress. Video can be recorded, student work can be shared, or written reflections can demonstrate a teacher’s abilities in a much safer way, providing the psychological safety necessary for growth.

**Goal setting and evidence collection provide disconfirming data, which leads to survival anxiety.**

In the goal setting process, teachers were asked to think carefully about their teaching and identify an area in which they would like to grow. They had to reflect and
identify some disconfirming data, or information that motivated them to change. The process of collecting evidence of progress towards that goal induced a feeling of survival anxiety in participants. Knowing that they would have to share their progress with peers and an administrator and have proof of it motivated them to take action. Nancy explained, “I don’t think there’s anything wrong with anxiety. A great amount of anxiety is not going to have anybody do better work. But you know, there’s that kind of band, I won’t say anxiety, but that band of competition which, you know, if done right, if managed well, you know, moves any group forward.” Though others did not feel “competition” was the best word to be used, they agreed that there was a “good pressure” involved in evidence collection when shared with peers.

*Goal setting and evidence collection provide opportunities for scanning and cognitive redefinition*

When teachers set goals and have to collect evidence of progress towards those goals, they are forced to scan their environment and their work for evidence. In efforts to achieve their goals, they also scan by talking to peers, going to the Internet, or reading to gain new knowledge. When setting a goal around mathematical fluency, Nathan scoured the Internet for research on best practices in teaching math facts as well as research on short and long-term memory. He went to his colleagues in the prior grade level and asked them how they taught math facts. He sought input from his study group when he did not see the progress he hoped for. These are all examples of scanning induced by the goal setting process. Nathan also experienced cognitive redefinition through evidence collection. What he thought would work simply did not. In his written reflection, Nathan shared,
I now believe that it took me too long to recognize that, while this might be a generally appropriate format for upper graders, it was too challenging for my students. I think now, that I paid greater attention to the literature than to the message my own students were sending me through their performance ... "This is beyond challenging, it is too hard for us." Thus, the key lesson I have learned, or at least been reminded of, is that close, empathetic attention to and reflection upon student work (and data culled there from) is an essential element in determining successful lesson content and pacing.

Nathan came to new understanding about his teaching and student learning through his evidence collection.

**Goal setting and evidence collection treat teachers like professionals.**

One explanation for why participants found this process motivating may be explained by Fairtlough's (2007) responsible autonomy. In such a setting, individuals or groups have autonomy to make decisions, but are also held accountable for those decisions. This accountability occurs through what Fairtlough describes as critique, when an external agency evaluates the autonomous unit. Doing so encourages that unit to act responsibility by holding it accountable for its actions. By setting goals and collecting evidence of progress towards those goals, teachers are held accountable for their own growth and are honored for exercising their judgment. They know that a group of peers and administrators will be expecting updates along the way as well as a final presentation of evidence. Again, their survival anxiety prevents them from coming to these meetings unprepared. They do not want to appear incompetent or ineffective in front of their peers and/or supervisor.

In addition to the pressure of being ready for meetings, seeing tangible growth in one’s teaching or growth in student results can also be motivating. In a traditional system of teacher evaluation where feedback can be vague and infrequent, it is difficult to truly
grasp how, or if, a teacher has improved. By clearly setting goals and collecting evidence of progress, concrete examples of growth can be seen. The growth becomes evident in student work samples, video of lessons, or even in the quality of a written self-reflection. The change is clearly visible and observable. There is a clear connection between effort and result, which may be lacking in a more traditional system, again serving to motivate participants.

**Frequency - an ongoing and iterative process.**

Most states require evaluations every two to three years and do not dictate how many evaluations occur during an evaluation year (Cohen, et al., 2008). In California, an administrator can observe a teacher for 20 minutes once during the year and still meet the requirements of California Education Code. This means a teacher can be judged on an observation of less than 0.1% of his or her actual teaching. It is not hard to imagine why the current system cannot be expected to change teachers’ behaviors and help them grow as professionals.

**An ongoing process increases psychological safety.**

The familiarity of an ongoing process coupled with larger amounts of available data increased the psychological safety that teachers felt. Nathan agreed, saying, “the more regularly it [evaluation] occurs and the more ... familiar you become with a process the less anxiety it can produce.” There is often anxiety around the unknown. With the infrequency of current evaluations, teachers often do not know what to expect and worry about the evaluation. As with anything, familiarity brings comfort. The more often something happens, the less threatening it becomes. In addition to increased familiarity with the system, the ongoing nature of the pilot provided much more data with which to assess
teaching and growth. When referring to the frequent administrative visits, Nancy stated, “It really feels good actually when an administrator really knows your teaching.” Through this process, teachers understood that a single lesson did not have to be perfect. Not only their lesson delivery, but also their planning, thinking and reflecting were valued in this pilot. This reduced the anxiety commonly associated with a scheduled administrative observation where teachers felt pressure to deliver the perfect lesson in the one interaction with the administrator. In this pilot, teachers had multiple venues to prove their value and worth as teachers, thus reducing their performance anxiety.

An ongoing process provides opportunities for disconfirming data and survival anxiety.

When thinking of teaching as a system of inputs and outputs, information about the outputs of a system can be fed back into the system. With a single observation, the feedback is also singular. Though a teacher may change after receiving feedback once, change is much more likely when the feedback is ongoing and supports a cycle of continuous learning. As Fairtlough (2007) would argue, a system that regularly loops feedback back into itself is much more responsive to its environment and constantly adapts as new information comes in. In the traditional system, a teacher who is struggling to motivate an inattentive student might receive a suggestion after an observation such as giving the child an object to fidget with during lesson delivery. But what if that does not work? What if the child plays with the toy and becomes even more distracted? Does the teacher need to wait another two to five years for a new suggestion? Though this is an exaggeration, a system that embraces ongoing dialogue about teaching supports continuous learning and collaboration amongst teachers as well as with administration.
This promotes double-loop learning where teachers can not only give and receive feedback on their teaching, but question their underlying assumptions and beliefs as well (Kauffman, 1980). The ongoing nature of this system provides many more opportunities for participant to be exposed to disconfirming data. With only one observation, the data may be discounted by explaining that it was just a bad day. Knowing that progress towards the goal will be revisited regularly and evidence of progress is required, teachers feel survival anxiety to make progress that otherwise might not be present in a single observation.

*An ongoing process provides opportunities for scanning and cognitive redefinition.*

The regular interactions with peers throughout this process over a longer period of time provides many more opportunities for teachers to scan for new information. Participants in this study shared that not only did they meet during designated study group meetings, but more informal meetings occurred between sessions. Teachers would pass each other in the hall and ask questions or sit together at lunch and brainstorm new ideas. Isabelle observed three other colleagues teaching mathematics and scanned their classrooms and teaching for ideas to support her progress towards her goal. This would not happen in a typical evaluation. The length of the study provided these opportunities for learning.

In addition to scanning, an ongoing process also increases the likelihood of cognitive redefinition. Being exposed to a new idea or way of thinking once may spark a teachers curiosity, but having to continually reflect on and discuss one’s progress increases the likelihood of actually changing one’s beliefs. Isabelle explained, “I’m taking that one goal and really putting into my everyday teaching.” Rather than trying something once for a
formal observation, she attempted to build thinking routines into her math on a daily basis and was rewarded for those efforts by developing a better understanding of how thinking routines are best applied to mathematics instruction.

*An ongoing process aligns with conceptions of teaching as a profession and art.*

The conceptions of teaching as a profession and teaching as art recognize the importance of teachers using their judgment and intuition. An ongoing and iterative process provides teachers with more opportunities to practice using their judgment and understand their intuition. Through conversations with peers and written reflections, teachers articulate how they used their judgment and intuition to make decisions about their teaching. Like any skill, judgment and intuition must be practiced, reflected on, and discussed in order to be developed. This raises the conversation to a different level of depth and complexity. This is double-loop learning. Teachers (and administrators) questioning their underlying assumptions and operating norms or standards is a much more complex task than comparing one’s self to a preexisting set of standards. The ongoing and iterative nature of the pilot aligned with the beliefs of teaching as a profession and teaching as an art in this way.

**Learnings, Habits, and Thinking Transcended this Process and Infiltrated other Areas of Participants Teaching**

One of the most exciting discoveries of this process was how participants' growth towards their goals was not limited to the goals themselves. Participants reported changes in other aspects of their teaching as well. The learning branched out beyond the scope of their professional growth plans. Nancy’s newly developed fluency with thinking routines in Black History Month seeped into her math curriculum. Cathy's lessons on persistence in
writing arched over into students persisting when sounding out unknown words. A compare and contrast method Loretta used to teach division was also found to be helpful in later mathematics unit. Beyond these explicit connections to teaching, there were also less direct benefits that developed outside of the evaluation process itself. Through the study groups and a changing relationship with an administrator, teachers developed a new support group. Teachers now had four more resources to go to when struggling with a lesson or other challenges in the classroom. They developed a habit of engaging peers in conversations and built comfort with those in their study group in particular. These systemic supports and changes in relationships may be of more value in the long term than the progress towards the specific goals in the professional growth plans.

**Connections to Research Questions**

**What do Teachers at Mountainview Unified School Consider to be Important in Evaluating Teachers?**

Teachers at Mountainview Unified School shared what they considered to be important in teacher evaluation both explicitly and implicitly. Explicitly, they stated:

- Teachers need to feel safe and supported in order to take risks
- An evaluation process should be ongoing and iterative
- A system of evaluation needs to be flexible in order to meet the needs of every teacher
- The process of evaluation should not be a heavy burden and detract from classroom activities
- An evaluation system should motivate people to engage in thinking about their practice
Teachers’ implicit beliefs were seen through their choices when developing the pilot and in their reflections:

- Teaching is more than lesson delivery. Planning, thinking and reflecting are important aspects of teaching that need to be valued
- Expectations need to be clear and efforts should be made to ensure common understanding of language.
- Opportunities to interact with peers are valued

**What do Teachers at MUS Believe are Appropriate Ways to Evaluate Teachers?**

Teachers’ beliefs about methods of evaluation were discovered both in the conversations that took place during interviews and group discussions as well as in decisions made about the pilot model of evaluation.

- Teachers should set goals and collect evidence towards progress of those goals
- The use of video tape and peer observation are appropriate
- State test scores should not be used when evaluating teachers
- Any other method of evaluation is appropriate so long as it can be defended.

With the exception of state test scores, the teachers in this study were open to just about any idea regarding evaluation, so long as they had some control in the process and could defend its value. This was seen when creating the list of potential evidence to be used in portfolios, and teachers agreed to adding “any defensible other” as an option.
After Participating in an Action Research Model to Develop a Teacher Evaluation System, How Satisfied are Teachers with the Co-Developed Teacher Evaluation Process?

Teachers were clearly satisfied with the co-developed evaluation system. Tracy exclaimed, “I think this is a great new idea ... I hope it takes off,” and, “In my 46 years in education it is the most productive evaluative process in which I have participated.” Each of the eight participants shared their satisfaction with the system in their final interviews. Because the conditions for learning and growth were created in a safe and supportive environment, they felt empowered and motivated to improve their practice.

What are the Steps, Challenges, and Lessons Learned in the Process of Teachers Co-Creating a Teacher Evaluation System?

Steps.

Though the steps are outlined in detail in Chapter Three, it is worthwhile to note key attributes that occurred in this process. The first step is understanding what teachers believe about teaching and evaluation. If the culture of the teachers and district focuses on teaching as labor and craft, the evaluation system is going to need to look much different than the model developed here. Districts in program improvement or using scripted curricula may need a much more structured system to meet the needs of program requirements and teachers’ desires. However, there is also an opportunity in this step to shape some of these beliefs by exposing teachers to research on evaluation. Teachers may not see the value in certain methods until being presented with research. This step is critical. Without sharing research, there is a risk that the end model will only reflect the
teachers’ potentially narrow views of teacher evaluation and may preclude ideas simply because teachers were not aware that certain methods were even possibilities.

After collecting this information, it is important to involve teachers in thinking and decision-making throughout the process. For the system to work, teachers must feel safe. Including them in the process helps ensure that they feel good about the process and product. This process is intended to flatten the traditional education hierarchy of administrator and teacher. By excluding teachers from decision-making, there is the risk that it will feel like another top-down directive from administration and teachers may hesitate to fully engage themselves in the evaluation.

Finally, reflection is a critical step in the process of co-creating a system of teacher evaluation. Despite the best intentions and collaborative nature of the planning process, the system may not work because of unforeseen circumstances. For this reason, time needs to be dedicated to checking in with the group and reflecting on how the process is working for the group. This should happen during the pilot as well as afterwards. In this way, the process itself is modeled after the goals of evaluation, growth and continuous improvement. It also signals to the teachers that taking risks and making mistakes is expected and encouraged. Challenges are seen as opportunities for learning and growth, refining the process along the way.

**Challenges.**

One challenge to the process is the scarcity of time and energy in the teaching profession. Teachers often feel stretched to their limit and taking on additional responsibilities can cause anxiety. Throughout the study I made every effort to be respectful of teachers’ time and was careful not to ask too much in regards to meeting time,
reading, and writing reflections. Even with these efforts, there were still times when participants were unable to complete work or attend meetings. My biggest struggle was in getting the participants to write reflections. Though I explicitly avoided adding work during key times such as report cards and parent conferences, unforeseen drains on time such as parent meetings, other district work, or personal illness made it difficult to get each participant to participate fully. One way to combat this was through clear communication with early notice of upcoming expectations. When meetings were coming up or writing was due, I gave much advance notice with reminders along the way. However, as seen in the final interviews, teachers did not feel that this system was burdensome. In fact, quite the opposite was true. Because the system empowered and motivated them, they saw the value of their efforts and considered them to be more than reasonable for the benefit they received.

A second challenge in this process is building a genuine climate of trust and safety where risk-taking is encouraged. Even though I clearly stated that this was the climate I hoped to achieve, an untenured teacher may still feel pressure to perform perfectly when I observe a lesson. In addition to the many ways the development of psychological safety was described above, avoiding any judgment of the lesson delivery can was critical to combatting this fear. By celebrating the thinking that occurs after a lesson is given, the teacher is reassured that mistakes are allowed and encouraged. As Moss Kanter (1983) claims, “We need to create conditions, even inside large organizations, that make it possible for individuals to get the power to experiment, to create, to develop, to test – to innovate” (p. 15).
Lessons learned.

There is a delicate balance between providing clarity, being precise with language, and developing an open and flexible system that promotes thinking and growth.

Throughout the action research cycle, participants demonstrated an eagerness for clarity, to the point of ensuring that individuals all had the same understanding of singular words such as “portfolio” or “evaluation.” The power of language and pitfalls associated with misinterpretation were obvious in interviews, group meetings, and individual assignments. In particular, the teachers in this study did not feel good about the word “evaluation” itself. Historical connotations of the word imply negative judgment and discipline. If the purpose of evaluations is seen as teacher growth, using the word “evaluation” inhibits risk taking and can set a negative tone to the process. If the word is to be used, the understanding of the word must be articulated and agreed upon to set the stage for openness and growth.

Similar hesitancy was observed around my use of “portfolio,” where teachers’ historical understanding of the meaning of portfolios led them to be biased against the idea. However, after developing a different and common understanding of the word, the participants felt it would be a valuable instrument. This happened on a number of occasions, leading to the important lesson of being clear with language and expectations. However, there is a delicate balance between achieving clarity and leaving room for thinking and independence. If the process is overly focused on clarifying my expectations the teacher is no longer responsible for thinking independently. There needs to be enough clarity to allow for common understandings of the process while maintaining enough
freedom for teachers to meaningfully engage in the process and exercise their own judgment. Much like Fairtlough’s (2007) responsible autonomy, individuals should have the ability to make decisions, but also need to be held responsible for those decisions.

One example of finding balance between clarity and flexibility was seen in the evidence section of the professional growth plans. I clarified the two items that teachers must have, but allowed for flexibility within those items. Teachers were asked to write a reflection and were given an expectation regarding its length, but the content was left open. However, it was not completely open. I provided teachers with a list of prompts that they could choose from, but also allowed them to respond in any other meaningful way that occurred to them. By giving sample prompts, it gave them an idea of the types of thinking I was looking for, but remained general enough for them to make it their own. Another example of this balance in evidence was present in having some items that were required and others that were not. Furthermore, independent thinking and creativity were encouraged by including a category named “defensible other” in which participants could include any imaginable piece of evidence so long as they could justify why it was appropriate.

Modeling was another key strategy in providing clarity coupled with flexibility. By providing samples of professional growth plans, teachers were provided with a scope for the work ahead of them as well as the level of thinking, reflection, and detail I was looking for. This relieved some anxiety related to the unknown nature of implementing a new system. Modeling also provided an opportunity to set high expectations. Similar to Vygotsky’s (1978) beliefs about children’s learning, the participants seemed to “rise to the intellectual life around them.” By sharing both my models as well as teachers’ individual
drafts of plans, teachers seemed motivated to take the work more seriously and work towards higher levels of engagement. Simple exposure to new ideas was a powerful force that shaped participants’ experiences. For example, although no participants included the use of video initially in their plan, once it was introduced in conversation to the groups, five of eight participants ended up including this new idea in their plans.

*A climate of trust and safety promotes vulnerability, openness and risk-taking.*

Teachers at MUS believe that the purpose of teacher evaluation is to help them grow professionally. Growing implies change. It is a universally accepted truth that change is often difficult for adults. A climate of trust and safety can promote these risk-taking behaviors that lead to change and growth. Historically, teachers at MUS were unclear about the purpose of evaluation. It often seemed like they were simply trying to check off a box or show the administrator what they thought s/he wanted to see.

One of the first keys to developing a climate of trust and safety is being explicit with the purpose of the evaluation. From the beginning of the study, I made the intentions of the evaluation system clear, to help them learn and grow as professionals. I was not trying to catch them at anything or rank teachers from good to bad. The system was meant to engage them in thinking about their profession and provide supports that would help them become even more skillful teachers. This was a huge shift from the typical power dynamic between administrator and teacher. Rather than judging, I was now supporting. My role shifted from identifying their weaknesses to helping them be successful. This was critical in developing a climate of trust and safety.

Knowing that teachers were not simply being judged on single observations also promoted a climate of trust and safety. Tracy shared, “If I knew that you were going to
come in to observe the [one] lesson I would plan something I knew I would be successful at and not take any risk.” The use of portfolios embraced a multiple measures approach that signaled to teachers that lesson delivery alone (particularly a singular lesson) is not the sole judgment of their value as a teacher. Though the focus of this system was not on judging, teachers felt comforted knowing that their peers and I could find merit in multiple aspects of their work.

*Discussions guided by research can help change teachers’ beliefs about evaluation.*

Teachers can have firmly held beliefs about evaluation rooted in their own experiences, but sharing research on the subject can open them up to new understandings and beliefs about evaluation. Participants reported changes and affirmations in their beliefs about teacher evaluation after reading the literature review, mentioning that perspectives were opened and feeling moved to reflect on experiences and desires regarding evaluation. Others noted a strengthening of existing beliefs, now knowing that they were confirmed with research. After reading the particular literature review used in this study, a teacher described how she now clearly understood why a small number of formal evaluations a year by an administrator does not lead to growth. This is an important insight when considering how to best create buy-in and motivation for a new tool. Findings from this study lead me to think, if teachers can be led to understand why the system makes sense, they will be more likely to embrace it.

*New ideas and ways of thinking sometimes have to be intentionally introduced.*

Great ideas can be ignored solely because people are unaware of them. The idea of using portfolios to collect evidence may have never arisen if teachers had not read the
literature review and been convinced of their value. Even when participants read about the value of using video to evaluate their teaching, one group unanimously adopted this strategy while the second did not. Because the idea of videos came up in the second group late in the conversation, it was not included when the first few participants discussed their plans. The lesson here is that when searching for great ideas, one must cast a wide net and expose the group to diverse ways of research-based thinking. Furthermore, if specific desirable activities have not made it to the discussion, the administrator can intentionally re-introduce the concept. However, care must be taken to avoid the impression that the administrator is directing the process, thus risking loss of psychological safety. Ideas should be presented in a unbiased manner so that participants can digest the idea without feeling manipulated.

Activities intended to encourage flexibility in thinking and openness to suggestion should be done collaboratively and informally prior to any written documentation.

When too much thinking and/or writing occurred prior to collaborative processes, participants were less likely to listen to suggestions with an open mind and change their thoughts and plans. Participants were more open to this disconfirming data when decisions about their plans had not already been made. This was seen clearly in the initial stages of writing professional growth plans, but was also seen in interviews where participants had already reflected in writing. Once a participant put thoughts to paper, they were less likely to be open to new ideas or change their opinions. When seeking truly open and flexible discussions, administrators should take care to abstain from writing activities prior to meeting.
Recommendations for Policy and Practice

The findings of this study suggest many implications for the future of teacher evaluation. In addition to specific logistical recommendations, efforts must be made to be clear with the purpose of evaluation and teachers should be involved in the development of new systems with more ongoing and frequent structures. Systems should be flexible to meet the needs of diverse teachers and settings, involving peer collaboration, goal setting, and collecting of evidence.

Logistical Outline of an Evaluation System

Though initially reluctant to put hard numbers to the flexible system that participants appreciated, they recommended a twelve-week cycle containing four formal meetings with opportunities for informal meetings and activities in between. The brevity of a twelve-week protocol would relieve any undue pressure to create works of art and compete with one another. However, a brief cycle would not provide the opportunities to experience all the components that lead to motivation and change. As Nancy stated, there simply would not be enough to reflect on.

The group felt that the process should start with a goal-setting meeting where participants develop their professional growth plans collaboratively. They would then have two mid-evaluation check-ins (one after four weeks and one after eight weeks) to share progress and brainstorm ideas. The final meeting would take place at the end of the twelve weeks with study group members sharing their portfolios with one another. Participants agreed that study groups of three or four members would be an ideal size for providing enough data while still feeling safe and supported. The group also agreed that
teachers should experience this cycle annually, as it is hard to argue that teachers only need to grow every other year.

**Work Collaborative With Teacher to Explicitly Define the Purpose of Evaluation**

Teachers want to know the purpose of an evaluation process. Are administrators trying to identify poor teachers? Are they simply jumping through an administrative hoop? By clearly defining the intent of an evaluation system, teachers can feel safe and supported. By including teachers in defining the purpose, the locus of control is shifted from the administrator to the teacher, empowering them to take charge of their own learning and motivating them to take the process seriously. As discussed above, by co-developing the system and explicitly defining the purpose of the evaluation as growth, they knew the system was meant to support them, not judge them. The intentions of the evaluation system were clear, to help them learn and grow as professionals. Our hopes were to engage in thinking about the profession and providing supports that would help them become even more skillful teachers. Shifting the dynamic from the judging to supporting is critical when trying to encourage learning and growth as well as developing a climate of trust and psychological safety. It signals that it is acceptable to be imperfect so long as one works to understand and improve those imperfections. Teachers reported feeling safe to take risks in this study. Nancy explained, “Having a support group of teachers made it easier to take risks.” The fact that teachers felt safe and were open to growth was critical to the success of the project. If teachers only wanted to get the evaluation done, the system used would be irrelevant.
**Involve Teachers in the Development of the Evaluation System**

The system would not have been as effective if the group had not participated in the action research process together. The process of reading the literature review and co-creating the pilot system led to a deeper understanding of why the system should work, serving to motivate teachers to engage in the process thoughtfully. Isabelle summarized the group's thinking, saying,

I’m not so sure how much buy-in it would’ve gotten had you gotten the eight of us together and then said, “Here are some final products; here’s what we did last year. What would you like to do this year?” I think it might’ve looked a little more – I don’t know, intense is the word that comes to mind – but because we were involved from day one in the building of it and what it might look like, and there was so much flexibility in-between, then that felt really comfortable. And then the staff just took it from there and excelled.

Being involved from the initial stages led to motivation and buy-in. They were invested in the success of the system from the very beginning. Though some guidelines can, and should, be given when developing systems at other schools or districts, efforts must be made to involve teachers in the process of creating the system. Without this step, there will be fear caused through lack of understanding of motives as well as the value of the system itself. The entire system can be sabotaged from the very beginning if teachers do not feel psychological safety.

**Teachers should Experience Evaluation through Collaboration with Peers**

As noted throughout this study, the power of peer collaboration is enormous. It helps lower anxiety, motivates others, exposes them to new ways of thinking, and benefits all of the individuals involved. Furthermore, peers are a resource that every school has easy access to. It is a largely untapped resource. Throughout this research, the value of
providing opportunities for peers to interact was seen. It was clearly a two-way relationship where those asking for help and those giving help benefited.

As schools look at redesigning their systems of evaluation, efforts must be made to create structures that force peers to interact throughout all phases of evaluation. The importance of peer collaboration and adult learning, often referred to as professional learning communities (PLCs), is shared by many educators. “Well-implemented professional learning communities are a powerful means of seamlessly blending teaching and professional learning in ways that produce complex, intelligent behavior in all teachers.” (Sparks, 2005, p. 156). Newmann and Wehlage add, “If schools want to enhance their organizational capacity to boost student learning, they should work on building a professional community that is characterized by shared purpose, collaborative activity, and collective responsibility among staff.” (1995, p. 37). Collaborating with peers can also lead to powerful normalizing behaviors and heightened expectations (Fenwick, 2001). Participants in this study tended to normalize the scope of their work with one another. There were no significant in outliers in the amount of effort put into the process as well as the quality of the work. By regularly sharing goals, progress and ideas with one another, participants came to develop and understand standards of quality in one another's work.

People can be naturally competitive, and when presented with high quality work of their peers, can be motivated to new levels of engagement with the evaluation process. When peers get together with the expectation of convincing one another why goals were or were not met, they are roused to action. No one wants to look bad in front of his or her peers.
Teachers Should Set Goals and Collect Multiple Forms of Evidence to Demonstrate Progress Towards those Goals

The process of goal setting and evidence collection was critical in focusing the work as well promoting reflection. Setting goals can help teachers begin with the end in mind and can provide structure to the use of portfolios (Brogan, 1995). Fenwick (2001) found that the use of professional growth plans developed a greater sense of commitment and focus among teachers. These results were replicated in this study as well. Having concrete goals and a plan to achieve them made growth realistic and achievable. Rather than simply crossing fingers and hoping teachers improve because they are being watched, goal setting and creating plans empowered them to make real change.

As discussed in the literature review, evaluations have no meaning unless there is reflection. “Teachers will not be able to increase their capacity for accurate self-assessment in the wake of isolated external value judgments,” (Feeney, 2007, p. 193). The process of reflection converts the traditionally passive system of teacher evaluation to an active system of learning and growth. The use of goals helps focus this reflection and evidence collection for portfolios helps ground the reflection in reality. This is the piece that truly engages teachers in thinking about and understanding their practice.

A System of Evaluation Should be Ongoing and Frequent

The current system of infrequent pre-announced administrative observations does not lead to teacher growth. With observations occurring only once or twice a year, the minimal amount of feedback is unlikely to change behavior. It can also frustrate teachers because it is only a snapshot of their teaching ability and does not represent the complex
thoughts and decision-making processes that occur throughout the year. All eight participants stated that evaluations should be more frequent and/or ongoing.

This increased frequency can lead to more psychological safety, which in turn leads to more risk-taking. Tracy shared, “If I knew that you were going to come in to observe the [one] lesson I would plan something I knew I would be successful at and not take any risk.” Knowing that a single lesson did not have to be perfect, teachers were empowered to take risks. When an evaluation relies on a single observation, there are huge disincentives to risk-taking. Risk-taking is needed to foster professional growth (Brogan, 1995). Not only does this structure inhibit risk-taking, but it also does not provide the ongoing feedback needed for change. As discussed earlier, in a responsible autonomy, critique needs to occur on a regular basis for the system to be effective.

**The System Must be Flexible**

Like our students, each teacher is unique. For a system of evaluation to be successful, it must be flexible enough for diverse teachers to find it valuable. A large part of that diversity can be found in offering choice. Rather than a one-size-fits-all approach, a menu of activities and flexibility in goal setting can provide a structure where teachers feel valued and connected to the evaluation. The challenge is in finding the appropriate balance between flexibility and focus. This can be accomplished by having a few core activities that are required in addition to choice activities. This balance could also be struck in goal setting, where a teacher might have the right to exercise choice in one goal while the administrator chooses a second. If this feels too “top-down,” the second choice might be from a predetermined list inclusive of school initiatives or other areas of focus. However,
providing choice is critical. When teachers choose, they become invested and involved in the process. Evaluation shifts from something done to them, to something done with them.

**Conclusion**

As teacher evaluation continues to be a hot topic in education, it is critical that we understand what its purpose is and how it can be leveraged to achieve teacher growth and thus gains in student achievement. As Bill Gates (2012) said in his recent piece in the *New York Times*, “Developing a systematic way to help teachers get better is the most powerful idea in education today.” In this study, we developed a systematic way to help teachers improve by providing psychological safety and motivational structures. When asked if she thought this model could help teachers outside of MUS, Tracy shared, “You don’t know until you give them a chance. Give them the trust and see what happens.” Ethan summarized,

> The schools that are most dynamic, most effective, most growing, are the ones where teachers are empowered, where you put the power into teachers and you support their passions and the things that they’re doing. And I think those are the ones that continue to grow, because you’ve got a learning organization happening where people are, you know, excited!

We need to trust teachers to be professional and provide them the support they need to learn and grow.

**Limitations**

While this study provides meaningful insights into the realm of teacher evaluation, motivation and improvement, there are limitations to consider.

When using the more traditional social science research definition of “generalizability,” it is clear that this study is not generalizable because of the unique setting of the school and experiences of the participants. Neither is representative of schools and teachers generally. However, the findings in this study are likely to be
transferrable across a variety of contexts. Particularly when the findings are understood through the lens of Schein’s (1995) theory, it is reasonable to assume the factors affecting teachers’ feelings of safety, motivations, and other beliefs are common among the general teaching population. Adults learn and grow in similar ways regardless of the context and have many of the same motivational factors. For these reasons, I argue that the findings are transferrable to other contexts.

My relationship with the participants was unique in that I am their direct supervisor and relatively new to the organization, being in the middle of my second year as Chief Academic Officer. There is a possibility that participants may have been affected by a desire to please me or say what they thought I wanted to hear. However, this reactivity was tempered by the fact that action research is a common practice at the school, and leading initiatives and making decisions in this manner is quite typical. Furthermore, prior to this study I had worked closely with the staff on developing social norms, one of which centers on having direct and respectful conversations. The school culture is one in which people are encouraged to speak freely. This was observed throughout the research with participants openly disagreeing with me (in a respectful manner) throughout the process, admitting a lack of knowledge and uncertainty, and taking risks.

Lack of Evidence of Definitive Impacts on Student Achievement

Although this study explored how a well-designed system of teacher evaluation can lead to teacher improvement, there was not an explicit instrument to specifically calculate and articulate how student achievement was influenced by participation in the pilot. Rationally, a system designed to help teachers learn and grow should positively impact student achievement, yet the study would carry much more weight if this could be said
with more confidence. Teacher responses anecdotally implied gains in student achievement and understanding but did not have the rigor of a more longitudinal and quantitative study. However, with any study of a complex system such as education, clearly proving causation between interventions and student achievement is difficult.

**Recommendations for Further Research**

As mentioned above, an important piece of continued research would focus on correlating co-created evaluation systems that are aligned with teacher beliefs with student achievement. The findings of this study and common sense suggest that this system would lead to increases in student achievement, but would be strengthened with quantitative analysis of measurable outcomes. The difficulty in this research would be identifying appropriate measures of student achievement. As noted in chapter two, the use of high stakes test scores in value-added assessment is fraught with inaccuracies. Also, because of the flexible nature of goal setting used in this study, development in a particular area such as creating better classroom discourse or infusing more thinking routines into the curriculum do not have directly measurable outcomes related to student achievement. However, by using agreed upon measures of student achievement, we could hope to prove correlational advantages that come from generally participating in the evaluation system. Furthermore, more research needs to be done on what we should actually be measuring. Is performance on a standardized test the best indicator of future success? Should we consider assessing traits of learners such as persistence, flexibility in thinking, and critical thinking?
Dissemination of the Findings

The most obvious dissemination of findings will occur at the local site. As with other decisions at the school, committee decisions and experiences are reported to the whole staff for orientation and review. The findings will also be shared at a local school board meeting. In efforts to share the results beyond the local level, I will present my findings to the Santa Barbara educational community at a county-wide curriculum council meeting and small district co-operative meetings. My most ambitious plans to share the results of my research include seeking publication of an executive summary in major trade magazines such as *Educational Leadership* and *Education Week* with hopes of turning the research into a book to be distributed through a publisher such as SAGE or ASCD.
Appendix A
Timeline of Events

Week of October 10th, 2011
  Recruit participants and collect consent forms

October 24th – November 2nd, 2011
  Conduct initial interviews

November 28th, 2011
  Distribute literature review and journal prompts

December 15th, 2011
  Conduct first group meeting, developing pilot system

January 4th, 2012
  Present and revise evaluation pilot system

January 5th – February 24th, 2012
  Implement pilot evaluation system - field observation notes

Week of February 6th, 2012
  Mid-evaluation check-ins

February 24th and February 27th, 2012
  End of evaluation portfolio sharing meetings. Final reflections due.

March 5th
  Final whole group reflection to complete the action research cycle

Week of March 5th
  Final Interviews
Appendix B
Whole Staff Pre-Survey Protocol

Background information (Read aloud before asking questions 1-2)

Darling-Hammond, Wise, and Pease argue that a thorough understanding of context (i.e. teachers’ beliefs about the nature of teaching) is critical in designing successful teacher evaluation. Under the conception of teaching as labor, teaching activities are routinized and prescribed. The evaluation system acts as a direct inspection of the work (Darling-Hammond, et al., 1983). When teaching is conceptualized as craft, teachers are expected to have techniques with generalized rules for application. In this system, evaluation is used to ensure that teachers have these prerequisite skills. In teaching as a profession, teachers are expected to have these prerequisite skills, but also to exercise judgment in their application. Teachers need theoretical knowledge as well as practical techniques. The evaluations process becomes more of a judgment on the ability of teachers to problem-solve. In teaching as art, teaching techniques and the ways they are applied may be novel and unconventional. There is a shift away from standardization where teachers have to incorporate their pedagogical knowledge and their own personal resources and intuition. In this conceptual framework, evaluation involves self-assessment and critical assessment by multiple people.

1. Which conception of teaching (labor, craft, profession, art) do you most closely align yourself with? Why?

2. Which conception of teaching (labor, craft, profession, art) do you think describes the beliefs of the teaching staff at MUS? Why?

3. Describe your past experiences with teacher evaluation
4. What did you value about these experiences?

5. How would you improve these experiences?

6. What is the purpose of teacher evaluation?

7. What is important to consider in a system of teacher evaluation?

8. What are appropriate ways of evaluating teachers?
Appendix C
Interview #1 Protocol

Background information (Read aloud before asking questions 1-2)

Darling-Hammond, Wise, and Pease argue that a thorough understanding of context (i.e. teachers’ beliefs about the nature of teaching) is critical in designing successful teacher evaluation. Under the conception of teaching as labor, teaching activities are routinized and prescribed. The evaluation system acts as a direct inspection of the work (Darling-Hammond, et al., 1983). When teaching is conceptualized as craft, teachers are expected to have techniques with generalized rules for application. In this system, evaluation is used to ensure that teachers have these prerequisite skills. In teaching as a profession, teachers are expected to have these prerequisite skills, but also to exercise judgment in their application. Teachers need theoretical knowledge as well as practical techniques. The evaluations process becomes more of a judgment on the ability of teachers to problem-solve. In teaching as art, teaching techniques and the ways they are applied may be novel and unconventional. There is a shift away from standardization where teachers have to incorporate their pedagogical knowledge and their own personal resources and intuition. In this conceptual framework, evaluation involves self-assessment and critical assessment by multiple people.

1. Which conception of teaching (labor, craft, profession, art) do you most closely align yourself with? Why?
2. Which conception of teaching (labor, craft, profession, art) do you think describes the beliefs of the teaching staff at MUS? Why?
3. Describe your past experiences with teacher evaluation?
4. What did you value about these experiences?

5. How would you improve these experiences?

6. What is the purpose of teacher evaluation?

7. How, if at all, should an evaluation system differentiate between teachers at various stages of their career or skill development?

8. On what should teachers be evaluated?

9. If an evaluation does two things really well, what would they be?

10. What are two things an evaluation should not do?
Appendix D
Journal Prompts – Literature Review

The prompts below ask you to share new understandings of teacher evaluation after reading the literature review, to connect your beliefs about the continuum of teaching to evaluation practices, and consider which practices might align best in the specific context of MUS.

1. How, if at all, have your views on best practices in teacher evaluation changed?
2. How does your conception of teaching (labor, craft, profession, art) inform your opinions about various practices in teacher evaluation?
3. Which practices do you think would support teacher learning at MUS?
Appendix E
Group Discussion Protocol

1. Which conception of teaching (labor, craft, profession, art) do you think describes the beliefs of the teaching staff at MUS? Why?

2. How does our conception of teaching (labor, craft, profession, art) inform your opinions about various practices in teacher evaluation?

3. What is the purpose of teacher evaluation at MUS?

4. Which practices do you think would support teacher learning at MUS?
Appendix F
Mid-Implementation Check-in Journal Prompt

1. What impact have your efforts toward fulfilling your professional growth plan had on student achievement?”
Appendix G
Final Interview

1. How many years have you taught?

2. If you were to tell the California Department of Education how to roll out a program like this to all public schools, what would you tell them?
   a. How often should study groups meet?
   b. How often should teachers participate in this cycle? How long should the cycles last?
   c. What “must do” activities would you include? (i.e. how much peer observation, written reflection, number of goals, etc.)

3. What is the ideal role of an administrator in this process?

4. How do we know this process is good for students?

5. If we were to assume that some standardized measure of student achievement is needed to fairly evaluate teachers, what might that look like?

6. The process we developed was a formative one. How might a summative evaluation intended to be used in high stakes decisions, such as giving teacher tenure, look different, if at all?

7. How did frequent interactions with peers affect how safe and supported you felt during the process?

8. What effect, if any did the act of goal setting and collecting evidence (portfolios) have on your learning?

9. Which conception of teaching (labor, craft, profession, art) do you think describes the beliefs of the teaching staff at MUS? Why? Has your belief changed since the initial interview?
10. What do you now believe the purpose of teacher evaluation is? Did it change throughout this process?

11. What did you value about the action research process used to develop the system of teacher evaluation? (follow-up – did being involved in the process lead to any difference in motivation or buy-in?)

12. What challenges did you encounter in the process?

13. Do you believe the final product will lead to growth as a teacher? Why or why not?

14. Do you believe the final product supports MUS’ conception of teaching as (insert labor, craft, profession, art)?
Appendix H
Draft 1 of Evaluation Pilot

Plan for professional growth - Mountainview Unified School

STEPS:

Preparing for the year

Reflect on areas that you would like to grow as a teacher and professional this year. Draft a minimum of one personal goal and one goal that supports the work of the strategic plan. Make sure the goals are specific, reasonable. Start to think about ways of documenting progress towards these goals (see menu of available options)

Pre-conference

Meet with the Chief Academic Officer and your study group to discuss goals and agree to a plan of documenting growth. Complete the professional growth plan form.

Gather Evidence

Start to collect agreed upon evidence.

Thinking/Reflecting mid-term discussion with study group

Meet at least once with your study group and discuss the following questions:

• How's it going?
• What's going well?
• What isn't working?
• What evidence have you started to collect?
• What's next?

Assemble portfolio

Gather evidence collected throughout the year and write a 1-2 page reflection on your progress towards the goals you set

Final presentation to study group and CAO

Meet with your study group and the CAO and discuss your portfolio
MENU OF EVIDENCE

Review the options below and think about what pieces would demonstrate progress towards accomplishing your goals. Options will be discussed and agreed to during the pre-conference, though additional evidence may be collected at any time.

*Probationary teachers participate in 3 formal observations in addition to agreed upon evidence

MUST DO’S

ADMINISTRATIVE OBSERVATIONS (ANNOUNCED/UNANNOUNCED)

Part of your evidence will include feedback from the CAO based on classroom visits. These can occur as drop-ins, by teacher invitation for a particular lesson, or other appropriate methods designed to provide opportunities to observe progress towards goals.

YEAR-END REFLECTION

Each teacher must submit a year-end reflection summarizing their growth for the year, lessons learned, etc. A length of 1-2 pages is recommended as a minimum. See the year-end reflection prompts for further guidance.

MAY DO’S

STUDENT WORK

Collect any samples of student work that demonstrate progress towards your goals. This could be multiple samples of high, medium and low work around a particular skill, sequential examples of a student's progress over time (i.e. documenting the writing process), video tape of group discussion or conference with a student, etc.

OBSERVING PEERS

Observe peers and document what you noticed and how it might impact your teaching.

BEING OBSERVED BY PEERS

Ask a peer to observe you and provide feedback. Briefly summarize their feedback and what you learned from the experience. Encourage the observer to avoid a “culture of nice” by providing honest and direct feedback.

VIDEO
Videotape yourself in any aspect of your work that demonstrates progress towards your goals.

PERSONAL REFLECTIONS

Write a personal reflection after trying something new, talking to your team, or at any other appropriate times throughout the year.

STUDENT ACHIEVEMENT DATA

Collect evidence of student growth on a particular topic or skill using teacher-created assessments, curriculum assessments, or work scored through rubrics. Pre and post assessments are suggested to demonstrate growth.

CONFERENCE ATTENDANCE

Include agendas and materials from any conferences you attended that are relevant to your goals.

PROFESSIONAL READING

Write a brief summary/reflection of any professional reading you've done that is related to your goals.

STUDENT/PARENT SURVEY

If you had a goal concentrating on feelings of students or parents, survey them on their feelings as appropriate.

SELF ASSESSMENT

Complete a self-assessment such as the Habits of Mind report card or any other instrument relevant to your goals (pre and post assessments can be particularly powerful in demonstrating growth).

LESSON PLANS

Include lesson plans related to your goals including items such as the sequence of activities, assessments, progress monitoring, student work samples, materials lists, etc.

DEFENSIBLE “OTHER”

Feel free to propose any other form of evidence that you can defend as appropriate to your goals.
YEAR-END REFLECTION QUESTION PROMPTS

Consider the questions below when reflecting on your growth for the year:

- What progress did you make towards your goals? Do you feel you achieved them? Why or why not?

- What were some key lessons you learned this year?

- What did you learn from your peers as part of this process?

- What piece of evidence were you most proud of? Why?

- In what ways did you contribute to the MUS community as a whole?

- Are there any opportunities to share what you learned with your colleagues?

- Did any of your work this year spark further interests of study you might consider for next year? What might you work towards in the future?
PROFESSIONAL GROWTH PLAN

Teacher ___________________________ Grade ________ Year __2011-12__

Goal 1:

Why is this goal relevant and important to your professional growth?

Narrative description of efforts that will be made towards achieving the goal:

Evidence to be collected:

Administrative commitments:
PROFESSIONAL GROWTH PLAN cont’d.

Teacher __________________________ Grade _____ Year __2011-12__

Goal 2:

Strategic Initiative # __________________________ Objective __________

Why is this goal relevant and important to your professional growth?

Narrative description of efforts that will be made towards achieving the goal:

Evidence to be collected:

Administrative commitments:

Date of Preconference ____________

Teacher Signature _____________________________ Date _____

Administrative Signature ___________________________ Date _____
PROFESSIONAL GROWTH PLAN (sample 1)

Teacher Bruski Grade 3rd Year 2011-12

Goal 1:

Develop fluency and accuracy in my student’s multiplication facts. 90% or greater of my students will be able to complete 100 multiplication facts (up to 10x10) with 98% accuracy in 4 minutes or less.

Why is this goal relevant and important to your professional growth?

School-wide conversations have indicated that many students still struggle with multiplication facts in the upper grades which makes learning more advanced concepts much more difficult. Our grade level as a whole recognizes a need to study how we teach multiplication and what strategies would support the student’s success in the upper grades.

Narrative description of efforts that will be made towards achieving the goal:

I will begin by reaching out to my peers at MUS, the math coach, and colleagues from past teaching experiences to gather strategies and materials that they have found to work in the past. I will explore the Internet for reputable and research-based strategies on improving fact fluency, reading a minimum of 3 articles. If possible, I would like to attend the fact fluency 1-day workshop put on by SBCEO in the spring for more ideas. Ideas I know I would like to explore are the use of student goal setting, having students track their own progress, systematically teaching the facts in a sequence to mastery, working closely with parents to emphasize the important of fact mastery and share games with them, and the use of music as a mnemonic device.

A large component of this goal is tracking student progress and collecting data. I will administer a pre-test to all students and track their progress regularly throughout the year, both in comprehensive timed tests and more specific “fact-family” clusters.

Evidence to be collected:

Student assessment data
Student progress tracking logs
Communication to parents (weekly letters, list of games, email, etc.)
Videotape of introductory lessons
½ page summary of key learning’s from colleagues, Internet, and workshops
Year-end reflection of what did and didn’t work (to be shared with my team)
Agenda from Fact Fluency Training
Administrative commitments:

4-6 drop-in visits during multiplication instruction (teacher to notify of best times)
PROFESSIONAL GROWTH PLAN cont’d.

Teacher ___________ Bruski ___________  Grade __ 5 __ Year __2011-12__

Goal 2:

Improve the “culture of thinking” in my classroom. Students will engage in high level (deep?) conversations around core literature evidenced by the questions they ask and the inferences and observations they make in writing, classroom discussions, and book club conversations.

Strategic Initiative: #1 Student Learning and Achievement_
Objective: Educational Framework - assess student work for evidence of critical thinking_

Why is this goal relevant and important to your professional growth?

4th and 5th grade have noted that more than ever it is difficult to get students to go deeper with the content and really push themselves. Although I’ve learned much at Harvard, I don’t think I’m getting as much bang for my buck with thinking routines because I haven’t developed as strong of a culture of thinking as I could.

Narrative description of efforts that will be made towards achieving the goal:

I will study Ritchhart’s Making Thinking Visible book and note a specific sequence of activities I’ll use to help build a culture of thinking. One idea I would like to explore is co-creating a rubric with students using the Habits of Mind and Thinking Dispositions that would allow them to self-assess their own writing as well as group conversations. This will build their awareness of their own thinking and give them the metacognitive background to better analyze their work through this lens. I’ve developed a strong dialogue with my study group of Harvard PZ and will float ideas to them as well for advice and reflection. I will also complete a WIDE World online course of study on cultures of thinking. My grade level team will be an important piece of thinking about this work together and refining I throughout the year.

Evidence to be collected:

Video of book club student conversations with dialogue tracked
Student writing coupled with their self-assessed rubric scores and rationales
Before and after writing samples or video samples from beginning of the year to end.
Email correspondence with Harvard PZ group
Feedback from peer observations
WIDE World course syllabus and completion certificate
Year-end reflection on big ideas – what worked and didn’t work
Timeline of activities/discussions that promoted the development of a culture of thinking in the room

**Administrative commitments:**

3 specific observations of this process (beginning, middle, and end) by teacher invitation.

**Other contributions to the MUS learning community:**

In addition to being a thought leader during grade level and whole staff PLCs, I will also contribute to School Site Council and am willing to serve as secretary. I’d like to explore ways to also formally or informally be more of a leader in the area of Visible Thinking. I’m willing to participate in any study groups or committees that would further explore this work and I’d love to help plan/present our in-service days.

Or:

Being new to the school and new to the grade level, I don’t want to overextend myself by joining too many projects. However, I have a lot of insights as a new teacher and interesting experiences from other settings that I could contribute to the thinking on new staff transitions that will happen in the new year.

**Date of Preconference ____________**

Teacher Signature ____________________________  Date _____

Administrative Signature ____________________________  Date _____
Appendix I
Professional Growth Plans

**Teacher:** Cathy  
**Grade:** K  
**Year:** 2011-2012

**Goal:**  
Students will become more persistent in their writing and begin to have an understanding of why writing is important.

**Why is this goal relevant to your professional growth?**  
Being new to Kindergarten, I am still getting used to all of the curriculum, so I can still grow in all areas. I chose to focus on writing since it connects closely with the current readers workshop unit and is an area that really interests me. Currently, the students are at a point where they are able to come up with ideas for their writing and use their sounds to put their ideas on paper. I’ve noticed that many students are becoming frustrated once they have completed their work because they are having trouble remembering what they wrote. In other words, their work has errors such as lack of spacing, messiness, etc. which makes it difficult for them to read or for their peers to read. I feel that I need to stress the importance of writing and come up with a checklist of sorts for the students to use to edit and revise their work. I want to develop and improve the students’ persistence in their work habits as well. Persistence connects to the school wide goal of incorporating the Habits of Mind into the classroom and is a vital trait for students to have to be effective writers.

**Narrative description of efforts that will be made towards achieving the goal:**  
Below are some ideas for lessons that will be taught throughout the unit:

- Why is writing important? Where in life is writing important? We’ll brainstorm ideas and show examples (recipes, street signs, directions, etc.) to support the idea.
- What types of things should our writing always have or always look like? Develop a checklist with student input for students to use when they are finished with their work and need to check it. This will kind of be like our class’ “Writing Non-Negotiables.” It will include things such as must have spaces, punctuation, pictures that match the writing, etc. We will post these on the wall and give students a personal copy for their reference.
- What does it mean to revise and edit? - How to use the checklist tool that we created together.
- What does it mean to persist/Why is this important to do? How does persisting connect to editing and revision?

We will look at each other’s work as examples in lessons when applicable and practice these skills whole class, with a partner, and individually.

I will collect student work and write down observations during work time as ways to track progress.
Evidence to be collected:
  - Student Work samples (Same students’ work taken throughout the unit to show growth)
  - Videotape of student/teacher conferences as we are discussing student work together
  - Lesson Plans
  - Student reflection on their work

Administrative commitments:
Observe during Writers Workshop lessons (dates and times to be determined by Mrs. Cloud)
PROFESSIONAL GROWTH PLAN

Teacher ___ Tracy _______________ Grade ___2___ Year ___2011-12___

Goal 1:

Develop student's understanding of Habits of Mind and ability to identify them in historical figures and themselves through a study of Black History Month while also building my own fluency with thinking routines

Why is this goal relevant and important to your professional growth?

I'd like to improve how students approach learning and think that increased awareness of habits of mind coupled with better fluency in thinking routines will improve student learning

Narrative description of efforts that will be made towards achieving the goal:

We will study historical African American figures and identify Habits of Mind that they demonstrated in their lives. This will include reading Ruby Bridges, studying Jackie Robinson and Martin Luther King to name a few. Through the use of thinking routines such as circle of viewpoints, perceive, know, care about, here now, there then, and 3,2,1, students will become more familiar with the Habits of Mind and recognize them in others and themselves. Students will also demonstrate critical thinking by examining a cause and effect chain of events – what might have happened if Ruby Bridges did not have persistence?

Evidence to be collected:

Students will create a “kidtionary” in which they identify key Habits of Mind in themselves Videotape conversations/interviews with children about their Habits of Mind Student work samples from thinking routines Final reflection

Administrative commitments:

Mr. Bruski will observe specific lessons 2-3 times throughout the unit (per teacher invitation).

Date of Preconference __1/5/12____

Teacher Signature ___________________________ Date _____

Administrative Signature ___________________________ Date _____
PROFESSIONAL GROWTH PLAN

Teacher:  Nathan  
Grade:  4th  
Year:  2011-12

Goal: Develop fluency in student addition and subtraction facts:
12 of 14 students (approx. 86%) will complete 50 addition and subtraction facts with 98% accuracy in a period of 2 minutes or less in 4 out of 5 attempts. (Resources consulted indicate that a reasonable speed goal for upper elementary addition/subtraction facts is 30 correct per minute/See ldonline.org: Recall_of_Addition_and_Subtraction_Facts).

What are the addition and subtraction facts?
Addition and subtraction facts can be defined as follows:
Addition: single digit addends with sums up to 18. For example, 7 + 6, 8 + 4, and 2 + 3 are addition facts, but 11 + 12 is not. This is because 11+ 12 has multi-digit addends and its sum is greater than 18.
Subtraction: inverse of addition facts as defined above, for example 13-7, 12- 8, and 5-2 are subtraction facts because they are inverses of addition facts. 23-11 is not a subtraction fact because its inverse does not fall within the definition of an addition fact.

Why is this goal is relevant to my professional growth?
I am currently teaching a leveled math group comprised of students for whom math represents a significant academic challenge. Automatic recall of addition and subtraction facts (fact fluency) is essential for success with higher level math skills. When students posses automatic recall they can quickly retrieve fact answers from long term memory without conscious effort or the need to rely on procedures such as counting on fingers, which can lead to errors. Additionally, automaticity frees working memory so that it can be devoted to problem solving, data interpretation, reasoning, and learning new concepts. Studies show that fact fluency level is a significant predictor of performance on standardized tests.

Narrative description of efforts that will be made toward achieving the goal:
*Because addition/subtraction fact fluency is ordinarily more strongly emphasized (and achieved) in lower grades, I will consult with lower grade MUS teachers, as well as MUS math coach, to gather strategies and materials which have proved successful in addition/subtraction fact fluency attainment.
*I will administer a pretest to students and establish systematic introduction of small sets of new facts using appropriate strategies (Resources consulted indicate that brief fluency practice sessions with limited, yet increasing, volume of material yields most productive results).
* As part of my assessment data compilation, I will engage students in goal setting and self-
tracking. These will be important components of this program because they can provide purpose and direction, help students establish priorities and stay motivated. Setting, tracking, and achieving goals also helps students recognize their strengths and specific needs for further development.

Evidence to be collected:
* Student assessment data
* Student goal setting and progress tracking logs
* Summary of key concepts from colleagues, internet, and other sources
* Cycle-end reflection including critical data and observations related to program’s strengths and areas of needed improvement.

Administrative commitment:

4-6 unannounced drop-in visits
PROFESSIONAL GROWTH PLAN

Teacher  ____Isabelle____  Grade  ____4____  Year  ____2011-12____

Goal 1:

I will seek to make students’ thinking in mathematics visible through the use of thinking routines and develop my own confidence and ability to implement thinking routines in mathematics.

Why is this goal relevant and important to your professional growth?

This goal is relevant because as a school we are trying to move forward with creating cultures of thinking and implementing what we learned at Harvard University. So far, these particular learnings have not been implemented thoroughly in mathematics. We know how important true understanding of mathematical concepts is, and I believe using thinking routines in math will help support this.

Narrative description of efforts that will be made towards achieving the goal:

I would like to get more information about how students think in mathematics by having them engage in thinking routines and writing to talk about their understanding in math and how they solve problems. I will attempt 3-5 thinking routines in the next unit and analyze their responses to see if I can better understand where their confusions lie so I can better intervene. Through this, I also hope that the students take on more of the ownership of the learning process rather than always being the one delivering the information. I will also observe a peer, and be observed by a peer, when teaching via a thinking routine in mathematics.

Evidence to be collected:

Student work from thinking routines
Documentation of what I learned from their responses and what I did differently based on what I learned
Notes from observing a peer teaching a thinking routine in math
Notes from being observed by a peer
Video of a lesson
Final reflection

Administrative commitments:

Mr. Bruski will observe specific lessons 2-3 times throughout the unit (per teacher invitation).
Mr. Bruski will "drop in" 3-5 times during the math block
Date of Preconference 1/5/12

Teacher Signature ___________________________  Date _____

Administrative Signature ______________________
Professional Growth Plan

Teacher: [Redacted]
Grade: 4th
Year: 2011-12

Goal 1:

Develop my use of Thinking Routines in math in order to deepen students' understanding of the content and strengthen their use of the thinking moves associated with a mathematical concept. Within a unit of study, I will incorporate 3-5 Thinking Routines and also have students complete a reflection piece of writing at the end of the unit.

Why is this goal relevant and important to your professional growth?

Since I teach the challenge math class at 4th grade, I am concerned with utilizing structures that provide for deeper understanding of the content, as well as extensions of the learning to more advanced concepts and real world situations. Additionally, as per the [Redacted] Strategic Plan, the strategic initiative of Student Learning and Achievement outlines 2 goals that address student thinking, in terms of a Culture of Thinking within the classroom as well as through the use of Visible Thinking strategies. Incorporating Thinking Routines in my math class would allow me to continue to develop a Culture of Thinking in my math class as well as utilize Visible Thinking strategies.

Narrative description of efforts that will be made towards achieving the goal:

I will draw from various resources, such as the Project Zero research and materials and another book entitled Math Tools, Grades 3-12: 64 Ways to Differentiate Instruction and Increase Student Engagement. I will also discuss and gather ideas from various other teachers on the staff as well as my team. In looking at additional ways I might be able to deepen students' understanding in math, an area I would like to explore is that of the International Baccalaureate schools' and how they design their math curriculum. So far, I have not been able to locate much in the way of Teaching for Understanding in relation to math, but I am interested in that as well, i.e. creating Throughlines in math. However, I believe IB schools design and structure curriculum with goals of deeper thinking and understanding, so perhaps I might be able to find resources that would aid me in this goal.
Evidence to be collected:

List of which Thinking Routines & End-of-Unit Reflection Piece will be incorporated within the math unit
Sample assignment sheet (or explanation of) each TR utilized
Student samples (written or pictures of posters) of their use of Thinking Routines
Videotape (or have a colleague observe) one TR utilized within the unit
Summary Reflection on the TR on the video OR
Notes from the post-discussion with colleague who observed
Samples of end of unit student reflections
Teacher reflection at end of unit that incorporated TRs on my perception of how well the TRs supported and deepened student thinking and understanding
If possible, notes from a discussion with an IB teacher (Could be a teacher of any grade, 3-8, and discussion would focus on math)

Administrative commitments:

___ Observation(s) during math instruction incorporating TRs (teacher to notify of best times, and if possible, observation of end of unit, culminating discussion following completion of student reflection pieces.

Date of Preconference: ______________________

Teacher Signature __________________________ Date ________________

Administrator Signature ____________________ Date ________________
PROFESSIONAL GROWTH PLAN

Teacher ____ Ethan ____ Grade 5 Year 2011-12

Goal 1:

I would like students to become better at describing their mathematical understandings through the use of thinking routines.

Why is this goal relevant and important to your professional growth?

This goal is relevant because as a school we are trying to move forward with creating cultures of thinking and implementing what we learned at Harvard University. So far, these particular learnings have not been implemented thoroughly in mathematics. We know how important true understanding of mathematical concepts is, and I believe using thinking routines in math will help support this.

Narrative description of efforts that will be made towards achieving the goal:

Many students say, “it’s just how I do it” or speak through numbers when I ask them how they solved a problem. I would like to get more information about how students think in mathematics by having them engage in thinking routines and writing to talk about their understanding in math and how they solve problems. I will attempt 3-5 thinking routines in the next unit and analyze their responses. For example, I might use same, same different with different methods of solving division so that students might see the different values that come with each method. I will also observe a peer, and be observed by a peer, when teaching via a thinking routine in mathematics and focus feedback by having 2-3 guiding questions to consider during observation. Peer observation will be important because when implementing a new thinking routine, I will probably be more focused on how the routine is going rather than what students are truly doing.

Evidence to be collected:

Student work/writing from thinking routines
Notes from observing a peer teaching a thinking routine in math
Notes from being observed by a peer
Video of a lesson
Final reflection

Administrative commitments:

Mr. Bruski will observe specific lessons 2-3 times throughout the unit (per teacher invitation).
Mr. Bruski will “drop in” 3-5 times during the math block
Date of Preconference __1/5/12____

Teacher Signature _____________________________ Date _____

Administrative Signature ____________
PROFESSIONAL GROWTH PLAN

Teacher ___ Nancy ___ Grade ___ 6 ___ Year ___ 2011-12 ___

Goal 1:

Develop student's understanding of Habits of Mind and ability to identify them in historical figures, characters in a novel and themselves through a study of Black History Month, Civil Rights and the novel, Roll of Thunder, Hear My Cry while also building my own fluency with thinking routines

Why is this goal relevant and important to your professional growth?
This will give me the opportunity to develop more thinking routines for this piece of literature and increase my fluency in using thinking routines. In addition, I will find more opportunities for my students and myself to develop Habits of the Mind in ourselves by identifying successful strategies used by successful people.

Narrative description of efforts that will be made towards achieving the goal:

We will study the novel, Roll of Thunder, Hear My Cry, the development of the civil rights movement in the United States from 1864 to 1964, and historical African American figures who were successful in leading the civil rights movement and identify Habits of Mind that they these successful leaders demonstrated in their lives. This will include reading Ruby Bridges, studying Thurgood Marshal and Martin Luther King to name a few. Through the use of thinking routines such as circle of viewpoints, perceive, know, care about, here now, there then, and 3,2,1, students will become more familiar with the Habits of Mind and recognize them in others and themselves. Students will also demonstrate critical thinking by examining a cause and effect chain of events – how might civil rights in America be different if Melba Beal had not persisted in her quest to integrate Central High School.

Evidence to be collected:

Students will identify key Habits of Mind in leaders of the civil rights movement
Students will identify successful strategies used by characters in Roll of Thunder, Hear My Cry, and reflect on the effective strategies.
After studying the ways Habits of the Mind are used successfully, students will write an essay documenting their discoveries, or write a letter to the person (Melba Beals for example) thanking them for their contributions to civil rights.
Student work samples from thinking routines
Final reflection

Administrative commitments:

Mr. Bruski will observe specific lessons 2-3 times throughout the unit (per teacher invitation).
Date of Preconference 1/9/12
Teacher Signature ___________________________  Date _____

Administrative Signature ______________________  Date _____
PROFESSIONAL GROWTH PLAN

Sarah      2nd-3rd Grade Strings      2011-2012

Goal 1: To employ additional strategies in the strings classroom to deliver specified technique instruction to individual students without sacrificing large amounts of instructional time at the expense of other students.

Why is this goal relevant and important to your professional growth?
As I continue to tweak the strings program each year, I’m recognizing that while students are leaving the program reading music and playing stringed instruments, there are some areas of technique that are suffering at the expense of other time spent on other skills. I need to set students up to succeed in future musical endeavors and specific technique is an important component of their success.

Narrative description of efforts that will be made towards achieving the goal:
I plan to focus on a specific class within each grade level as my test subjects. I will isolate the specific technique needs for each student and keep track of my attention to those needs each week. I will seek out advice from fellow string teachers (Ron Zecher- BTSA mentor) regarding effective technique-teaching strategies.

Evidence to be collected:
Observing of peers- I plan to make use of my allotted BTSA days to go and observe Ron Zecher and his colleagues in the SBSD teaching strings

Being observed by peers- Ron will observe me teaching and provide feedback

Student work- I plan to videotape each student playing their instrument in the first and last weeks of the project, as well as once or twice in between in order to track progress.

Administrative commitments:
Drop-ins during specified classes
Appendix J
Pre-survey Responses

Pre-survey response 1

1. which conception – I actually align more with teaching as a science – sort of somewhere between profession and art. Clearly there are base line knowledge, skills, and dispositions that are necessary to be an effective teacher, yet with the complex nature of the classroom and the disparate needs, learning styles, and challenges of any classroom of students, there is an art of requiring sensitivity, instruction, intuition and varying interpretations that greatly influences how learning takes place on a given day.

2. conception of MUS – I would say, as a whole, we skew towards profession. It is a group that values constant updating in the theoretical arena and takes on the challenge of translating those theories into practice.

3. Past experiences in eval – Actually, I haven’t really every experiences effective teacher evaluation. At best, I’ve received a banal written evaluation based on a single observation. I’ve had pre-conferences that gave me hope for an evaluation based on goals and interests, but these were never followed up in the process. Recently, there has been no evaluation whatsoever.

4. What did you value – It was a nice boost to read positive comments and have some form of feedback rather than teaching in a black hole

5. How would you improve – Clearly evaluation at its best is a conversation. It’s a formative process that involves peers as well as administrators in a process of reflection, goal setting, establishing a focus, observed application in practice, more reflection with feedback and keeping on in a continuous process.

6. Purpose – the bottom line purpose of teacher evaluation is increased success of students. As such, it needs to be a process that truly engages the teacher in active research and reflection into his/her own practice in an effort to add to and have the teacher’s knowledge, skills, and dispositions as they apply to the building, enacting and refining a classroom culture that supports the success of the students.

7. Important to consider – It’s formative. Empowerment, peer interaction, ongoing process of growth, supportive, research-centered environment

8. Appropriate ways – actually, the best model I’ve seen is the one used for National Board Certification

Pre-survey response 2

1. Which conception - I consider myself a master teacher. My work in the classroom is always at the professional level and most often practiced as an art. As a public school teacher, I believe that I have the responsibility for making sure that district curricular expectations and state standards are addressed in the curriculum. How I go about delivering that instruction and making decisions regarding the program for my class and individual students raises the teaching practice to the artful level.
2. Conception of MUS – I think the conception of teaching varies at MUS. I listen carefully to discussion in whole staff faculty meetings, committee meetings, and lunch table discussions. Based on my observations and day to day work with others, I believe the majority of teachers function at a professional level and many veteran teachers practice teaching as an art.

3. Past experiences –
   a. I believe the best teacher evaluation I have had at MUS was with Dr. Jack Smith, as a previous principal and superintendent. His management style gave teachers freedom to design their own classroom objectives and to monitor and work to meet those goals. He met with teachers two to three times during the year to discuss progress toward the goals and showed interest and involvement in the process. One did not have the sense that he was going about the business of evaluation but that he was actively involved in teacher growth. He showed genuine care and concern about the teaching and learning process at MUS.
   b. I have also had with two subsequent principals the standard fall and spring teacher/principal conference where we discussed students and program and set a goal for the year. Both administrators observed a lesson in the classroom and wrote a review at the end of the year.
   c. My most recent experience included meeting briefly with the administrator before a formal lesson observation. We then had a meeting where a response form was presented by the administrator.

4. Value - With (a) above, I valued being able to set individual goals and being given the resources and support to meet those goals. Dr. Smith believed strongly in giving teachers the opportunity to visit other schools, attend conferences, and/or have consultants come to the school. He allowed teachers to engage in goal setting and to own the process.

5. Improve process – Teachers need to be actively involved and engaged in the process of evaluation.

6. Purpose – To assure that students are receiving the highest level of instruction. To provide opportunity for learning and growth by the teacher.

7. Important to consider – The individual strengths and level of experience of the teachers. A one-size fits all approach will not always provide the most growth and engagement for all. There needs to be room for teacher input into the manner of evaluation.

8. See (a) above – assuming that the teachers are at the professional level. It goes without saying that teachers who are not functioning at a high level need more frequent and consistent observation and consultation with the administrator.

Pre-survey response 3

1. Conception of teaching – Combination of profession and art. Aspects of both provide a framework within which to operate with encouragement to be innovative.

2. Conception at MUS - Craft, profession and a bit of art. The expectation of MUS administration has shaped the teaching based on how teachers are evaluated, i.e. what is valued by the administration. The culture at MUS is that when a new
teacher comes onto a team, that teacher is “trained” by the team to continue with their past practices. Innovation and change are rarely (if at all) encouraged). That perpetuates a closed system which can become less effective rather than more.

3. Past experiences – Through the years evaluations have ranges from one sentence (“Loryn is doing well in all areas”) to a narrative paragraph. To a checklist – are based on traditional assembly line values.

4. Value – not so much except being recognized for doing well and not getting in trouble for some unknown reason

5. Improve – big job – that would involved retraining school systems, board, administrators, etc.

6. Purpose – to weed out the poor teachers, to keep an ongoing record. I’ve never felt I learned anything significant

7. Consider – the purpose, the criteria and the tools

8. Appropriate – I’m not sure anyone can answer that since few people agree what a good teacher is. Maybe that’s the starting point.

Pre-Survey Response 4

1. conception – I would most closely align myself with the art of teaching, because it is the combination of personal experiences, interactions and intuitions and pedagogical knowledge that allows a teacher to connect to each student. This combination of skills allows you to build rapport and trust, meet each child where they are, and be able to act and react to in response to each individual student’s learning style.

2. Conception of MUS – I also believe art is the conception of teaching that describes MUS. This is a staff of such experienced teachers that as a group, you are able to look at big picture ideas. The focus is on higher level thinking strategies, metacognition, thinking dispositions, which allows each teacher to use their own intuition while cooperating and building upon their own ideas.

3. Past experience – Because I have been a temporary for my entire career, I have been formally observed three times a year. This means I have been involved with using and reflecting on a variety of teacher evaluation tools. Last year, I worked closely with my principal to rework the evaluation tool for Oceanside School District.

4. Value - I value the tools and conversations that allow me to receive specific constructive feedback. I also value honesty and suggestions to improve my teaching practice.

5. Improve – The only experience that I have had the need improvement are ones where evaluation is a “satisfactory/not satisfactory.” As a relatively new teacher this does not allow me to grow in any way.

6. Purpose – the purpose of the evaluation is tool is to provide constructive feedback with specific objective evidence and suggestions for improvement. It should allow for a teacher to strive to meet new goals and should be a team effort between the observer and the teacher.
7. Consider – it is important to consider that each person takes feedback differently; many take it personally. I think it is important that the evaluation system allows for open conversation to celebrate successes and work towards common goals.

8. Appropriate – I think being supportive and clear in your purpose and focus is effective. I think having common goals and an atmosphere where you are working together as a team, rather than employer/employee. From an evaluator’s perspective, I believe it is valuable and appropriate to observe a planned formal lesson, but also drop by informally to see the typical environment of the classroom.
Appendix K
Coding Categories and Sub-Categories

Beliefs About 'Teacher Evaluation'

- Teacher Growth
- Motivation
- Accountability
- High Stakes Evaluation
- Student Achievement
- Lack of competitiveness
- Teacher Expectations
- Trust
- Fear of taking risks
- Safety

Understanding the framework as you see yourself
- Labor
- Craft
- Profession
- Combo
- Flexibility
- Processes
- Evidence Collection
- Goal Setting

Evaluation Practices
- Frequency
- Past Experiences
- Desires
- Personal Focus
- School Focus

Multiple Measures
- Portfolios
- Professional growth plans
- Alternative assessments
- Peer Observation
- Administrative role
- Reflection

Faults with current assessments
- Not measuring what is important
- Teaching to test
- Competitiveness

Value Added Measures
- Peer observation
- Administrative role
- Reflection

Frequency
- Past Experiences
- Desires

Goal Setting
- Personal Focus
- School Focus

Peer Observation
- Administrative role
- Reflection

Video
- Student Work
- Reflection

Observations
- Dialogue

Appendix K
Coding Categories and Sub-Categories
References


Ediger, M. (2000). How should teachers be assessed?


Sawchuck, S. (2009c). Peer review undergoing revitalization: An old concept that calls for teachers to assess their own is gaining traction as evaluation comes under the spotlight. *Education Week.*


