2017

A case study of the design and implementation of a new teacher evaluation system in a northern Michigan school district

Scott Moore

Follow this and additional works at: http://commons.emich.edu/theses

Part of the Education Commons

Recommended Citation

http://commons.emich.edu/theses/760
A Case Study of the Design and Implementation of a New Teacher Evaluation System in a
Northern Michigan School District

by

Scott Moore

Dissertation

Submitted to the College of Educational Leadership
Eastern Michigan University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
Educational Leadership and Counseling
Concentration in Educational Leadership

Dissertation Committee:
Ronald Williamson, EdD, Chair
Ella Burton, EdD
Diane Fox, EdD
William J. Price, PhD

September 25, 2017
Ypsilanti, Michigan
Dedication

This dissertation is dedicated to all of the people whom I have had the opportunity to serve in the past and will have the privilege of serving in the future. I am grateful, humbled, and honored to have been given that opportunity. I am hopeful that God positions me to help as many people as possible. I am thankful for his enduring love.

I thank my parents, Jeff and Gail Moore, for always supporting me. For your steadfast and countless acts of love, I could never fully return all that you both have done for me. I thank you for your unyielding support and I love you.

I have been showered with tremendous encouragement and support; through those gifts I have been empowered to finish the race. My grandparents, John and Edith Moore along with Gerald and Katherine Wilson, have blessed me beyond measure with their example and inspiration.

My friends and family, Susan L. Barnhart, David William Beauchamp, Betty Michelle Fisher, Robert Fisher, Kristy Grabow, Eric Hendricks, Nancy Hendrick, Randy Hendrick, Michael Anthony Lake, Derek Moore, Meredith Moore, and Charles J. Negro, have each provided me with countless exhibitions of support that was needed to complete this specific process. I am eternally grateful for all that each of you have done.

I pray that my children, Andrew and Hannah Moore, value the honor of serving others and the integrity associated with doing your best in all that you do. I am also hopeful that they respect the benefits of education and the contentment of maximizing your potential. My love for both of you motivates me to want to do my best.
Acknowledgments

I acknowledge and thank my dissertation committee and the faculty of Eastern Michigan University; their support, caring nature, commitment, and dedication are examples I hope to emulate. I sincerely thank Dr. Ronald Williamson, my dissertation chair and someone I look to as a mentor. His work ethic, thoughtful approach, integrity, caring, and skill have served as examples I hope to replicate in my own ventures of service. From my first experiences at Eastern Michigan University, he was there to support and guide me, keep me on track, and frame suggestions masterfully. Without his leadership and counsel, I am not confident that I would have been able to grow as I have.

I also thank Dr. Ella Burton for taking time one summer afternoon to simply ask how I was doing, as she could see I was under some stress. She took time to care, and I will always remember that. Dr. Burton played a pivotal role in the construction of this study and simultaneously gave me the confidence necessary to tackle the dissertation. I extend my gratitude to Dr. William Price for his counsel, skill, and proficiency. Dr. Price’s expertise in structuring this study cannot be overlooked, as he provided critical insight into what was necessary to construct a case study of this nature. I recognize how fortunate I am to have worked and collaborated with Dr. Diane Fox. Her interest in my topic made the journey rewarding and her insight was truly valued.

I acknowledge Dr. Christine Beardsley. Her willingness to share the truth about this process made it doable and not as intimidating. Your example was critical! Your support will always be appreciated. I acknowledge all of my teachers, friends, and family who have cared about me and encouraged me. Your impact on me will forever shape the landscape of my future, and I thank you.
Abstract

The purpose of this study was to identify how one district attempted to meet the need for greater accountability in the teacher evaluation process as mandated by law and to describe strategies that were employed by the district to ensure the integration of student achievement data. This research highlighted how a teacher evaluation instrument evolved over a 5-year period from 2011 to the end of the 2014–2015 school year.

This qualitative, single-case study comprised examination of interview responses of 17 stakeholders including a former superintendent, technology director, two board of education members, three former principals, and ten teacher-leaders. Open-ended interview questions led to an objective understanding of the process of change and compliance with Michigan legislation, PA 102 (2011). Findings revealed a consensus of the meaning interpreted by the personal experience of the stakeholders about the integration of Danielson’s (1996) teacher evaluation rubric with a system that considered multiple measures of student achievement data.

Multiple themes emerged in the analysis of stakeholders’ perceptions of the efforts in one Michigan school district to comply with a state-mandated teacher evaluation process. Paramount among the themes was that trust needed to be involved in the process of teacher evaluation for improvement to transpire. Principals’ leadership and collaboration among staff, union representatives, central office, and policy-makers was critical to successful implementation of the modified teacher evaluation process. Ownership felt by all stakeholders extended confidence for addressing continuing change and assured sustainability of the process.
Table of Contents

Dedication ................................................................................................................................. ii

Acknowledgments .................................................................................................................... iii

Abstract .................................................................................................................................... iv

List of Tables ........................................................................................................................... ix

Chapter One: Introduction to the Study .................................................................................... 1

   Background ......................................................................................................................... 1

   Problem Statement .......................................................................................................... 2

   The Purpose of This Study ............................................................................................. 3

   A Conceptual Framework for This Study ........................................................................ 3

   The Design for This Study .............................................................................................. 5

   Research Questions. The study was guided by the following research questions: ... 9

   Significance of the Study ................................................................................................. 9

   Summary ............................................................................................................................ 12

   Organization of the Study ............................................................................................... 13

Chapter Two: Literature Review ............................................................................................ 14

   Overview of the Review of Literature ............................................................................. 14

   Perspectives on Approaches to Education Reform ....................................................... 15

   Educator Perspectives on Achievement Data .............................................................. 20

   History of Teacher Evaluation ......................................................................................... 25
Appendix D: Administrator and Former Superintendent Interview Protocol...... 201

Appendix E: Board Member Interview Protocol.................................................. 202

Appendix F: Technology Director Interview Protocol ......................................... 203
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Study Participants by Role and Gender</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>Participants by Role, Pseudonym, and Perceived Adopter Category</td>
<td>71</td>
</tr>
</tbody>
</table>
Chapter One: Introduction to the Study

Background

Constructing and integrating a teacher evaluation instrument that addresses learning in the context of student achievement data and teacher performance is a trying task. Ahmadi (2013) noted that the work to mine effective student data for evaluation requires testing and analysis for a district to ensure that the measures implemented in defining teacher effectiveness are relevant and accurate, as perceived by the educators themselves. Coulter’s (2013) study cited teachers who believed the Danielson system provided useful tools for teacher development but failed to accurately introduce a fair means of quantifying student data results into the evaluation of teachers. Darling-Hammond (2013) reported that anxiety related to determining who is and who is not an effective teacher is one of the most pressing topics in education. Strickland (2014) conducted a case study of teachers’ struggles to integrate the Danielson model of evaluation with effective practice.

According to Clofelter (2006), accountability in respect to student achievement data is an important topic that continues to drive policymakers. Legislation has dictated that educators must demonstrate to society a return on the investment into student learning. The rapid approach to teacher evaluation taken by lawmakers across the country to institute accountability has failed to align criteria that is consistent with correlating effective instruction and student achievement results. The mixed methods study of Ramirez, Clouse, Davis, and White (2014) examined teacher evaluation and the ensuing frustration that educators encounter without a clearly defined system that leads to effective practice. The frustrations that teachers encounter appears to be immense in respect to identifying true measures of success for their students and interfacing that data with the process of
professional development as intended in the evaluation process initially created by Danielson (1996).

**Problem Statement**

Federal legislation prescribing accountability for school systems, such as No Child Left Behind (NCLB, 2001) and Race to the Top (2009), have been further extended at the state level. In Michigan, legislation regarding educator evaluations is compiled in Public Act 102 (2011), which comprises MCL 380.1249 and MCL 380.1250, and subsequent revisions and amendments.

PA 102 (2011) mandates that teachers who are rated as *ineffective* for three consecutive years must lose their job. The most effective staff will be retained, and educators who are not deemed as *effective* as their colleagues will be displaced, virtually eliminating seniority as the major factor in staff retention. Enforcing these standards has proven trying, as districts continue to grapple with aligning evaluation tools with student achievement data. According to the Mackinac Center for Public Policy (2014), many districts find the task of quantifying effectiveness so challenging that they have ignored PA 102 (2011) and utilized seniority for layoffs.

A comprehensive measure for teacher evaluation that includes student data has not been established by the State of Michigan. Each individual school district has been required to construct an evaluation system that comprises a qualitative instructional rubric with quantitative student achievement data results. The absence of a uniform measure to ascertain teacher effectiveness persists despite current laws that require teachers to be evaluated on student performance data as the primary determinant of employment. Each school district in Michigan has confronted the challenge of conforming to Public Act 102 by devising
measures of evaluation and creating their own unique systems of evaluation to quantify teacher performance.

The Michigan legislature has not dictated what student data to use when evaluating teachers. The lack of clarity leaves school districts, and their administrative staffs with little guidance for meeting the requirements of PA 102 (2011). In an effort to ascertain what identifies teachers as effective, a quantitative analysis of student achievement data must be included in the process of teacher retention as dictated by public statute. The intent of PA 102 (2011) was to link student achievement data with individual teacher performance. Because education is also inclusive of qualitative observation, the need to identify instruments that accurately reconcile qualitative observation data conducted by administrators with quantitative student achievement results is critical.

The Purpose of This Study

The purpose of this study was to identify how one district attempted to meet the need for greater accountability in the teacher evaluation process as mandated by law and to describe strategies that were employed by the district to ensure the integration of student achievement data.

A Conceptual Framework for This Study

Berlin (1953) contended that the importance of developing a strong conceptual framework hinges on the notion of capturing something real. In the past decade, few legislative mandates have had more real impact on public education in Michigan than laws concerning teacher evaluation.

The construct of change. This study was based on the concept that organizations that enact reforms are inherently exposed to a change process that is highly personalized, as
outlined by Hall and Hord (2015). Gall, Gall, and Borg (2005) described the inherent challenges relative to a district’s capacity to conform to legislative mandates of a rapidly evolving law. A basis for the present study can be found in the work of Rogers and Shoemaker (1971) that highlights the responsiveness of individuals when an organization confronts a change. Individuals can be identified as innovators, leaders, early majority, late majority and resisters/laggards. Through an explanation of the process undertaken by the school district, it was possible to identify individual roles as outlined by Rogers and Shoemaker.

Further, the framework of this study was derived from the principles outlined by Hall and Hord (2015), who claimed that change that is meaningful requires time and persistence. They also held that individuals will experience organizational change through a variety of stages stemming from their interpersonal needs and that administrative support is critical to this process. Hord, Stiegelbauer, Hall, and George (2006) stated that developing a mass of support is key to not just converting staff to buy into the change initiative but, more important, to ensure that students benefit from the proposed changes.

**Sense-making and interpretation.** As Choo (2006) determined earlier, prior knowledge and experiences was essential to this study for establishing a foundational structure necessary to consider new information, and developing the framework of the study. Pajeres (1992) believed that past knowledge and the integration of new information leads to new concepts and schema. This parlays into the linkage of concepts and eventually the ability to make predictions.

Spillane (1999) recognized that individual beliefs and opinions contribute to the assimilation of sense-making. As indicated by Hord et al. (2006), the change process is
ongoing and should not be characterized as an event. It is an extremely personal experience that requires the development and nurturing of feelings and skills. Consideration should be given to the people involved in change first and to the innovation afterwards. These lessons are more imperative in the challenge of a rigorous investigation and scrutiny of acquired knowledge in one’s own district to ensure that emotion is distinguished from knowledge.

Choo (2006) determined that in the context of sense-making during the course of social interaction, the need for individual stakeholder input is compelling. Hill (2001) conveyed that policy interpretation takes place in an individual’s mind. In the present study, the responses of stakeholders in the school district regarding issues relevant to PA 102 (2011) were examined at a micro and macro level.

**The Design for This Study**

This study examined how one northern Michigan school district addressed the challenges inherent with the enactment of Public Act 102 (2011). The district’s efforts to comply with PA 102 (2011) constitute the present case study. The research described how the district prepared for the implementation of this legislation and revealed the strategies implemented to build, nurture, and sustain the modified teacher evaluation system in the district through the spring of 2016.

An analyses of the districts’ efforts to conform to statute Public Act 102 yielded a thick description of qualitative data. The school district gathered data in the summer of 2010 via the InFORMED data results rubric. The measure of effective teaching adhered to the modified Danielson rubric that was produced in the work of Charlotte Danielson’s *Enhancing Professional Practice: A Framework for Teaching* (1996).
With specific emphasis on the State of Michigan since the inception of PA 102 (2011), a focal point of this research centered on the process used by a single school district in northern Michigan to develop the current evaluation instrument and how the tool has evolved since its inception. This study regarding teacher evaluation in its natural context provided emic and etic perspectives of the complex task of aligning a qualitative measure of teacher evaluation to reconcile with quantitative student achievement data as mandated by the State of Michigan. According to Kottak (2006), an emic perspective is held by those who have a local insight into the social phenomenon; an etic vantage point is derived from examining the phenomenon through a scientific lens, recognizing that the members within might be too closely involved to offer a totally unbiased account.

Striving to achieve an objective understanding of the process of change and compliance with PA 102 (2011), this research highlighted how the teacher evaluation instrument evolved over a 5-year period. The school district partnered with Zimco, Inc., of Frankenmuth, MI, to construct an automated mechanism to quantitatively evaluate teachers while simultaneously evaluating teacher performance with the qualitative rubric created by Charlotte Danielson (1996), which serves as a catalyst for teacher improvement. Teachers believed Danielson’s four domains relative to perceived teacher efficacy were relevant, according to Doerr’s (2012) quantitative study.

An overview of the case study approach used in this qualitative single-case study. Creswell (2014) wrote, “Case studies are a strategy of inquiry in which the researcher explores in depth a program, event, activity, process, or one or more individuals” (p. 14). Yin (2003) said that case studies can be utilized to explain the impact of a contemporary phenomenon that exists in the context of real-life circumstances. Krathwohl (2009) and Yin
(2009) agreed that a single case study isolates the true nature of the implications inherently inflicted by the phenomenon being studied. As is the case with this study, an understanding of the emerging legislative implications of teacher evaluation will emerge in the case study of the phenomenon involving the process, practice, and perspectives of stakeholders in a selected school district.

The case study approach has gained in acceptance over time. Eisenhart (1989) pointed out that case studies highlight or validate research models for a specific organization, wherein people and organizations are so diverse and unique in their nature. Dyer and Wilkins (1991) claimed that even one case might be worth studying if it leads to uncovering new theoretical perspectives and questioning previous thoughts held to be self-evident. Neuman (1994) cited how case studies develop the link that is necessary between teaching and research, ultimately leading to research-led instruction practices. Miles and Huberman (1994) believed that studies of this nature develop into a richness in research that could parlay into intimate understanding of the topics investigated.

Yin (2009) believed that a research case study might be the most appropriate methodological tool for a particular research inquiry if the objective is to conduct in depth contextual analysis of a specific phenomenon or situations that are limited in number. Jennings (1996) noted that there is no one systematic and uniform single process for addressing educational cases.

This approach to research and study is acceptable, according to Lundberg (2001), as the diversity that exists amongst educators, students, and institutions is immense. Bodgan and Bilken (2003) contended that an appreciation of the complexities and real-life application that ensue when conducting qualitative case study research will lead to the researcher gaining
unique insight into the experiences of the participants who lived in the context of the phenomenon.

Stratch, Pavel, and Everett (2008) wrote about the benefits of case study research as teaching tools, with the aim of bringing reality into the classroom. The authors pointed to Harvard Law School’s introduction of case studies in 1870 and the proliferation of that method of instruction in medicine, law, business, and management (p. 199).

**Data gathering.** A combination of information sources, termed triangulation or confirmatory, and perspectives gathered from in-depth interviews of stakeholders increased authenticity and accuracy of the data gathered in this case study. Field-based inquiry was heavily integrated into the study. Notes and video/audio recordings served as data, as well as open-ended interviews with the former superintendent, technology director, two board members, three principals, and ten teacher-leaders.

The notion that communication from the organization is critical to the advancement of any change was evaluated using the concerns-based adoption model (CBAM) developed by the American Institute for Research to evaluate the respondents’ stages of concern as the new process of evaluation was rolled out relative to communication and the levels of use once the instruments were integrated.

Documents and media analysis were reviewed with the objective of garnering an accurate assessment of the evolution of the evaluation phenomenon in the district and the impact and influence of the legislation. Teacher evaluation and student achievement data were gathered and analyzed using the modified Danielson rubric and InFORMED from 2011–2016 through STAGES, a division of K-12 Evaluation Solutions®, a web-based system designed to automate staff evaluations.
STAGES was designed in 1999 by Saginaw Valley State University through a U.S. Federal Teacher Quality Grant. STAGES has since been utilized by districts across the United States as a database to evaluate teachers, administrators, and support staff. This software integrates student growth and achievement data for an individual educator’s evaluation. InFORMED was designed by the school district and currently incorporates local and national achievement results. This tool was designed to be a customizable web-based instrument that allows users to choose which data pieces to integrate into the evaluation process.

A review of artifacts such as planning documents, board updates, media reports, and personal documents of key stakeholders were included representing how the district worked to create InFORMED and the modified Danielson rubric. A review of self-report measures such as self-evaluation tools and reflective data from principals and teachers were woven into the research.

**Research Questions.** The study was guided by the following research questions:

1. How did one school district prepare for the implementation of Public Act 102?
2. What strategies were employed to construct, and sustain support for the modified teacher evaluation?
3. What did the district learn from the modified Danielson rubric and the InFORMED data results rubric?

**Significance of the Study**

A plan for compliance. PA 102 (2011) mandates that teacher effectiveness must be quantified for the purpose of retention and the continuation of teaching staff employment. Toward that end, this study will help to identify instruments that accurately reconcile
qualitative observation data with quantitative student achievement results. Understanding the phenomenon of change and interpreting the value of the plan for compliance, educators will be more aptly equipped to navigate the challenges of the PA 102 (2011) legislation.

Danielson (2011) claimed, “Educators need to create procedures that yield valid and reliable results…they must be valid, measurable, reliable and defensible” (p. 6). In Michigan, this has yet to be done despite mandated legislation. Tienken (2011) indicated that the government is working on initiatives for which they do not have a target. The depth of stakeholder perspectives in the present study will lend credibility and validity to a plan for teacher evaluation and provide direction for other school districts at various stages of compliance with state legislation.

**Benefits of a qualitative approach.** A qualitative approach to explore evaluation of educational practice aligned with quantifiable accountability of student achievement provides an added dimension to previous research that has relied heavily on statistical findings. Tienken’s (2011) study, based on the idea of Goodhart’s Law, as described by Lewis (2004), found that relative statistical regularity will tend to collapse when external pressure is placed on data for control purposes. Campbell (1976) validated this as well by finding that the more quantitative pressures that are placed on social indicators, the more they will become corrupt. A study that converges data including the grades teachers issue, the local assessments created by a school district, and a nationally normed assessment such as NWEA to determine if there is a true connection to sound educational practice as outlined in the Danielson Framework could prove to mitigate the issues Goodhart described.

**Impact of professional development.** This research may be useful in applying the principles to teaching and improvement of student learning as those found in the work of
Feureborn and Sarin (2011) about adopting intervention strategies. The authors proved that professional development is essential to successfully establishing systematic change. They explained that any systematic change can derail progress. In this study, the process of exploration and professional development could provide insight into how to implement and adopt the Danielson framework if it is found to have an impact on student success as defined by local stakeholders in a rural Michigan district.

**Teacher accountability and retention.** Political, human, and social factors are colliding with a wide variety of information and ideals in respect to evaluating the effectiveness of educators. According to Smith (2014), the Michigan House Fiscal and Education Committee estimated that implementing new evaluation systems could have cost school districts 42 million dollars. Opfer and Pedder (2011) reported that professional educators generally agree that evaluation should be predicated on professional development and improvement.

Bates (2011) cited Eric A. Hanushek, an economist from the Hoover Institute, who said that the idea of firing your way to good teaching seems to be driving legislators in the evaluation process. The premise of Hanushek’s work focused on firing the bottom 8% of teachers; however, practicality in determining who is in the bottom 8% provides many obstacles. Schlechty (2011) highlighted the idea of *accountabilism*, that those not producing results be *eaten* or fired. Schlechty (2011) found that this concept discourages intrinsic value for educators and would be absent of the value of development for professionals. The results of the present study may support the work of Pan (2009), who believed that quantifying what is good teaching and correlating the results to an identified system of student-achievement data points would serve as the spring-board to improved instructional development.
**A training model.** Haviland, Turley, and Shin (2011) found that program assessment should consist of collecting data, reviewing data, and building a plan to improve. The political implications related to using the gathered information to hold teachers accountable are intense in the absence of accurate information needed to improve and the lack of clear data related to student achievement. In this study, as in the work of Haviland et al., allowing opportunities that are organic in nature to emerge during the evaluation process can serve as training that results in improved program assessment and buy-in of faculty.

**Summary**

This qualitative case study describes the experience of one school district in northern Michigan as they developed a process to comply with PA 102 (2011), State of Michigan legislation regarding teacher evaluation and its relationship to student achievement. The state mandate could be considered the most sweeping legislative educational piece since the federal No Child Left Behind Act (NCLB, 2001). Torres (2004) contended that enactment of NCLB marked the inception of the most dramatic effort since the Civil Rights Movement of the 1960s to depart from state control of education. According to Torres, the motivation for enacting such strong initiatives was done to ensure that measured instruction and learning is equally provided to students regardless of their assigned sub-group and to reduce performance gaps among students, eliminate incoherent curriculum and staff, and eradicate systemic complacency.
Organization of the Study

The State of Michigan enacted PA 102 (2011) to integrate state control of the educational systems it delivers. The background of the Michigan legislation, problem statement, and purpose of this study were discussed in Chapter One, along with the conceptual framework, study design, research questions, and statements of significance of the study.

Chapter Two comprises a comprehensive review of literature relevant to the history of teacher evaluation, student achievement data, educators’ perspectives on student achievement, teacher evaluation in Michigan, and the piloting process for teacher evaluation systems. Methods for the conduct of the study are discussed in Chapter Three, followed by the findings of this study in Chapter Four. Conclusions, implications, and recommendations for further study complete this work in Chapter Five.
Chapter Two: Literature Review

Overview of the Review of Literature

Gall, et al. (2005) contended that literature review serves multiple purposes, but that answering questions that arise throughout the course of educational practice is one of the primary purposes behind conducting a review of existing literature. Further, “Reviewing the literature will help the researcher generate ideas for framing new questions to investigate and for designing the methodology of the study” (p. 28). Rudestam and Newton (2001) noted that reviewing literature makes connections that bridge previous work that was conducted with the aims of the proposed topic of study.

With the intent of constructing meaning from the experience of one school district’s efforts to comply with mandated state legislation, various resources were searched for appropriate literary contributions to the study. Sources included online searches of key words and subjects through databases such as Google Scholar, ProQuest, and ERIC and research relative to teacher evaluation, student achievement, and stakeholder insight.

This study examined how one northern Michigan rural school district attempted to conform to the parameters outlined in Public Act 102 (2011) in reference to teacher evaluation. The changing landscape of federal and state statutes have required school districts to focus on the most effective policies and practices to evaluate teachers and increase academic outcomes for students.

Topics in this chapter include a review of perspectives on education reform and student achievement data followed by the history of teacher evaluation and legislative efforts toward that end in Michigan, including survey data, discussion of the piloting process of four specific observation tools for teacher evaluation and the struggle to assess and integrate
student achievement as it relates to teacher evaluation. The chapter concludes with key findings of pilot programs and recommendations of the Michigan Council for Educator Effectiveness and a discussion of the role of change and innovation in the school district’s process to integrate student achievement and teacher evaluation.

**Perspectives on Approaches to Education Reform**

*A return on investment.* Yeh (2010) conducted a study of the cost-effectiveness of 22 various approaches to education reform. He found that after eliminating variables, the rapid assessments of children for educational purposes garnered the best return on investment. As such, it is no surprise that policy-makers are insisting that standardized assessments provide the framework for evaluation criteria. However, these assessments continue to lack widespread uniformity. Strom (2011) reported that more than 14,000 distinct school systems exist in our country. With so many different perspectives at the national level, there may be too many variables extant to produce true data on what constitutes success.

Walters (2008) explained that society demands a return on investment from the industry of education. Walters isolated factors that are inherent as variables in respect to student achievement on standardized tests and applied economic models to the study. Walters said that society expects the return on the investment in public education to yield tangible results, considering the amount of resources allocated to the public good. He contended that districts are stewards of significant resources and the demand for maximizing the value and utility of these resources must be quantified.

Walters (2008) identified and defined efficiency as a qualitative factor in three various and distinct forms: technical efficiency, allocative efficiency, and x-efficiency.
Technical efficiency is shown in public officials being stewards of public funds, considering a variety of options, and being trusted to choose most wisely for the community they serve. Allocative efficiency is predicated on simply going with the least costly option as a means of selecting how to expend public funds. X-efficiency speaks to the notion that effects not readily or easily observed, such as vision, motivation, personal incentives, and leadership have a significant role in contributing to efficient practices.

Walters applied a useful study about where education in Michigan is directed in reference to accountability. From the perspective of parents and politicians, MEAP/M-STEP, MME, and the ACT seems to be essential components of the standards for which educators are to be held accountable as a significant portion of their evaluation. However, it is not possible at this time to tie individual student achievement data for state-issued testing in an equitable manner to individual educators in a timely fashion for evaluations due to the schedule of testing and the unknown factors relative to the changing process at the state level.

**Quantifiers of effectiveness.** Secretary of Education Arne Duncan (2009) claimed, “50 different states doing their own thing doesn’t quite make sense” (p. 3). The quantifiers of identifying which schools and teachers are effective are too wide in scope to define, according to Secretary Duncan. NCLB and Race to the Top have had a considerable impact on what the most significant factors are in respect to determining if a school is effective. Proficiency rates of 100% in math and language arts were expected for all individual students by 2014. Schools not meeting these parameters were to have sanctions placed on them.

The National Assessment of Educational Progress (NAEP) is the measuring stick for accurate proficiency ratings. Applegate, Applegate, McGeehan, and Pinto (2009) conducted an analysis of eight states and their testing system and did not find that there was a positive
relationship to NAEP. Kingsbury, Olson, Cronin, Hauser, and Houser (2003) found that state
standards and assessments are so distinct in defining proficiency that one test may have a
considerably different interpretation on student, teacher, and school performance than
another standardized national test. Cronin (2012) claimed that Northwest Evaluation
Association [NWEA] has accounted for a study that predicts student proficiency on the
MEAP in math and reading based on the student’s NWEA score. This information could
prove valuable as it could lend itself to validating each of the assessments interdependently.

NCLB (2001) legislation allows states autonomy to identify what is considered
proficient. Peterson and Hess (2008) conducted an assessment of states that are giving
accurate reflections of proficiency. Their study found that only three states were given a
grade of A. Most states had set such a low standard of proficiency that people became
disillusioned with what was identified as success. As a result, states applied for waivers from
the requirements of the federal Race to the Top (2009) program, such as the condition of
approval which raised state testing standards in respect to proficiency rates. Proficiency rates
for assessments had schools scoring 50% proficient or less and yet exceeding the state
average at this rate. As such, it is a safe assumption that there was considerable discrepancy
when graduation rates average around 80% for most districts in Michigan.

Bridgeland, Dilulio, Streeter, and Munson (2008) indicated that, according to parental
perceptions, underperforming schools are half as likely to inform parents of student success
rates as are high achieving schools. Johnson, Arumi, and Otto (2006) echoed that finding by
perception data collected that showed that 40% of parents surveyed did not feel that having a
high school diploma ensured that graduates had the capacity to read, write, or perform basic
arithmetic.
Hunt (2008) and Johnson (2003) both found that inflated grades led to condoned mediocrity, prevented identification of the true needs of learners, and ignored accountability for teachers and schools. Although grades are important to teachers and students as indicators of success, the government and parents do not necessarily always concur with those sentiments. Local common assessments with defined curricular standards seem to be a plausible solution to mitigate the potential of fraud. For this reason, it is not surprising that according to the *Educator Evaluations and Effectiveness in Michigan* (2013) survey 60.8% of district respondents utilize local assessments in their evaluation tools.

In Malikow’s (2006) review of the work of Darling-Hammond (2005), he found that where professional development is the focus, schools are dedicated to learning for the whole student in every cognitive and affective area. Schools committed to professional development maintain the following ten points among stakeholders: mutual self-interest, mutual respect and trust, shared decision-making, clearly stated objectives, manageable objectives, manageable agendas, accomplishable goals, commitment of leadership, fiscal support, long-term commitment, and information sharing and communication.

Economists Murnane and Cohen (1986) conducted research that provided information specific to the relevance of incentives in labor. They found that when people believe goals are achievable and desirable, incentives are effective. The issue with educators appears to stem from the intricacies with isolating the scope to which educators are able to impact student learning.

**Educators charged with responsibility for student learning.** Society seems to have shifted from putting the onus of learning on the student to the responsibility falling equally, if not even greater, to the teacher. Qian, Youngs, and Frank (2014) conducted a
study among school districts in Michigan and Indiana that indicated that even mentor teachers feel a tremendous amount of responsibility for the learning of the students in their mentees classes. Their study yielded insight into the drive of educators to ensure that learning of students, as opposed to what the teacher delivers, is the pinnacle objective. Measuring the outcomes from teaching to learning is the challenge that the districts of Michigan are charged with answering. Ainley and Patrick (2006) conducted a study embedded in psychological framework that indicates that even self-regulated learners need more than one item to accrue an accurate assessment of learning. Odden and Fermanich (2004) suggested that a multi-level educational mode to study teaching effectiveness and its relationship to student achievement tests is pivotal to understanding the effects of the variables and student characteristics that impact learning affirmation.

Gratz (2011) indicated that schools should determine factors that indicate teacher performance, identify outcomes of those factors, and the goals that should be met as a result. According to Cox (2011), teachers relate to learning from master teachers, which suggests an emphasis upon professional development as a primary factor in teacher evaluation. Teachers yearn for an effective measure of individual accountability that provides venues for improvement. Sherman (2006) claimed that teacher effects and teacher effectiveness are two distinct concepts that have considerably different meanings. In other words, teachers impact students, but how effectively they are leading them to success is different. However, as Darling-Hammond (2010) found in her empirical study, it is incredibly complex to assess teacher effects by imposing value-added measures that are often difficult to distinguish. Skinner (2010) cited that educators, unions, and administrators are united in their resolve that
reinventing educator evaluation and searching for ways to improve means of compensation for teachers is critical to improving student success.

Klein (2011) cited that the World Economic Forum ranked the United States as 48th in the realm of student performance on standardized math tests. Klein claims that this is due to the fact that politicians feel the need to assess teachers with value-added measures designed to identify a teacher’s impact on student test scores. Hanushek (2011) conducted longitudinal quantitative analyses of the impact of a teacher whose students score in the 60th percentile earning $5,000 more annually than teachers whose students score below this threshold. According to Kane, Tyler, and Wooten (2007), society is searching for an answer to why students are not competing on a global scale when, according to a study of 14 large school districts across America, 98% of teachers are deemed as satisfactory. The bewilderment of public policy officials is how $10.6 billion dollars is allocated to education at the state level in Michigan, yet children are not performing at the level of children in other countries. This provides further rationale as to the urgency to identify accurate measures to assess our teachers.

**Educator Perspectives on Achievement Data**

With the introduction of $1.2 billion dollars spread over five years to states that adhered to and adopted legislation related to Race to the Top, in 2010 the federal government persuaded states to rapidly find ways to compensate educators based on student achievement. Despite only two states being awarded the initial grant, many states, including Michigan, are now legislatively bound to identify programs that relate student achievement to teacher effectiveness and compensation. After reviewing the Project on Incentives in Teaching (POINT) in Nashville, Tennessee, Gratz (2011) surmised that, like Glazerman and Seifullah
(2010), the Teacher Advancement Program (TAP) in Chicago, Illinois, showed there was no correlation to teachers being motivated to improve practice by tying monetary rewards to student achievement. From a teacher and school perspective, success can be found through identifiable factors that lead to students learning at the students’ individual and respective level.

Darling-Hammond (2006) indicated that training and having highly effective teachers accessible to all students is the most important solution to the issues related to education. This is consistent with the feelings of educators, but identifying the measures and means to assess success continues to be a challenge. Schools across the country are subject to major discrepancies in funding and accessibility to highly skilled educators. Schools with lower socio-economic groups tend to be targeted as under-performing, and teachers find that variables such as family background often present obstacles too difficult to overcome.

Rockhind (2007) found that many teachers are willing to work toward student improvement, but do not see the utility and value of standardized tests. Nichols and Berliner (2007) noted that standardized tests often provide a mechanism to compare schools in a specific area, but can, according to teachers, lead to undermining the educational learning process. Further, standardized tests in Michigan are only used to assess roughly one-third of the instructors in a district. Based on the outcomes of standardized assessments, schools could receive sanctions for poor performance (United States Department of Education, 2002). NCLB extended into the credentialing of staff and insisted that instructors pass competency of content requirements and also be highly qualified in the subject they teach.

Grades as the assessment of student success. The parties who are impacted by teaching and learning are the students themselves, the school staff, the parents, and the
government as a conduit of society. Each of these groups has differing perspectives on what would constitute a successful learning experience. Student success can be interpreted differently by various stakeholders and benefactors of the educational process.

Strom (2011) believed that students are the most important stakeholders. His contention is warranted and would dictate consideration into student needs. Schoeffel, Kuriloff, and Steenwyk (2011) conducted a study wherein students defined educational success. Findings showed that students believed that grade point average (GPA) was the most critical factor in determining if success was achieved. Students felt that a balance between grades and the relationships they develop are critical to learning. The team also cited that the GPA for students who can claim that one adult at school is invested in their education is considerably higher than students who reported no relationships with adults. Dweck (2008) surveyed students and overwhelmingly found that students with higher GPAs believed that effort ignites ability and translates into success.

Students, as opposed to legislators, schools, and parents, appear to have a difference of opinion in respect to the value and utility that standardized tests have in the learning process. In a study conducted in Florida by Armstead, Bessell, and Plaza (2010) students generally felt that teachers exhausted way too much of their energies and efforts preparing for the Florida comprehensive assessment test [FCAT] that all students in Florida must take in high school. Students indicated that the onus placed on these tests detracted from learning, and when surveyed, they felt that preparing for the FCAT did not prepare students for life after high school. Armstead, Bessell and Plaza (2010) went back to the students and found that small learning groups and access to exciting instruction would help to prepare for ease of
entry into society. Students also believe that cyber skills are integral to determining success (Rosen, 2010).

The autonomy in how grades have been distributed has traditionally been at the discretion of the instructor. The variances that exist in grading are far too wide to explore. However, an understanding that grades are viewed as a key determinant of success is paramount to understanding teacher, student, and parent quantifiers of achievement. Marzano (2000) asked, “Why would anyone want to change current grading practices? The answer is simple: grades are so imprecise that they are almost meaningless” (p. 1.). As measures of specific standards, this proves to be true as grades have so much variance in the medium in which they are delivered. Grades are not consistent among teachers according to O’Connor (2011), but as Patterson (2003) contended, grading students fairly does not mean grading them equally. Brookhart (2004) indicated that grades are meant to serve multiple purposes and when you attempt to use one measurement for numerous purposes, it becomes difficult to make effective conclusions.

Bailey and McTighe (1996) conveyed that the primary purpose of grades is to communicate success to students, parents, school administrators, post-secondary institutions, and employers. Teachers use grades as their interpretation of success. As such, grades are the assessment of a teacher’s interpretation about whether a child was successful. There are many downfalls to utilizing grading as the sole measure of success. Gathercoal (2004) concluded that it is inappropriate to utilize grades for extrinsic purposes, because it detracts from the intrinsic value of learning. However, it appears impossible to eliminate grading as an ingredient for determining student success. Wiggins (1999) found that grading is only objective when looking at calculations, factual information, and spelling; otherwise, it is
human subjectivity that determines grades. This subjectivity will prove impossible to isolate, as too many variables are involved in the practice of issuing grades. From the student’s vantage point, the inclusion of teacher-issued grades appears critical to the quantitative analysis of teacher effectiveness, and should be included. Notably, an entire domain of the Danielson rubric is dedicated to classroom environment. Classroom environment is reflective of human interactions, and grades serve as a depiction of effort and academics.

However, the ultimate objective of the teacher is to ensure that students are successful. Defining success and to what degree appears to be the focal point of determining what impact the teacher potentially had in the success or lack of success for student learning. While some indicators of student achievement are universal, identifying measures of success that are inclusive of invested interest groups is the first step to delineating whether an educator has achieved the expectations set forth by stakeholders.

**A consistent approach.** The system as a whole is moving towards a more uniform approach to providing instructional services and analyzing the effectiveness. Linda Darling-Hammond (2006) studied educational systems across the globe and concluded that the United States are among the most unequal in providing well-trained instructors and resources in an equitable manner among industrialized nations. A cohesive system must be integrated to reverse these trends according to Darling-Hammond. A study conducted in South Carolina by Darling-Hammond (2004) indicated that having well-qualified and trained instructors can account for an 84% variance in student achievement.

The need to define the effectiveness of an educator and simultaneously provide development for educators in the process appears to be of the highest priority to the field of education. The intent of a thoughtfully designed assessment, such as the InFORMED rubric
would ensure consistent measures to define successful practice across the educational industry. A consistent method of accounting for student success will prove pivotal, particularly where failing schools are concerned. Darling-Hammond (2003) found that the majority of failing schools are in urban settings, and that achievement gaps can be reduced if well-trained educators are in place.

**History of Teacher Evaluation**

A portrait of the phenomenon of teacher evaluation emerged in a review of the history and evolution of teacher evaluation from the 1700s to the present. Tracy (1995) reported that clergy were the first to provide organized education in North America. They had the most formal academic training and were also perceived to be the community’s servants. Burke and Kerey (2005) contended that communities began hiring individual supervisors or supervisory committees cloaked with unlimited power to hire and fire educational staff, though no defined pedagogical expertise existed. Marzano, Frontier, and Livingston (2011) credited a rising industrial base at the beginning of the 1800s for establishing a call for a principal teacher to assume the role and responsibilities of an educator evaluator. Further, the sprawling urban movement of the mid-1800s led to creation of more organized school systems and the recognition that the assessment of a teacher’s effectiveness was a complex endeavor that required training and expertise. *The 1845 Annual Report of the Superintendent of Common Schools of the State of New York* (Young, 1845) acknowledged that method of instruction and teacher efficacy is the primary objective of an educator.

The work of the one of the foremost writers and innovators in education history begin to take shape in the early 1900s. John Dewey (1938) introduced methods of instruction that remain benchmarks today, such as student-centered education, real-life applicability to
instruction, differentiated instruction, and cross-curricular integration of content. According to Marzano et al. (2011), the approaches to education espoused by Dewey were saturated in the idea that democracy and qualitative observation of instruction was the key to educational delivery reaching its potential. At the same time Dewey was influencing the educational landscape (McDermott, 1981), Frederick Taylor (1911) approached learning in a scientific manner from an industrialist vantage point that was predicated on the efficiency of identifying the one most effective method to complete any task. Taylor noted that specific tasks, such as the hiring of personnel, professional development, and the division of labor could be identified as best practice. This measurement approach to instruction and defining student success by measuring student outcomes was beginning to have influence on the process of educator evaluation. By 1929, the book *Public School Administrator* was released by Cubberley (1929). His work advocated the inclusion of tangible data measurements to assess the effectiveness of teachers and schools and how to provide specific feedback designed to improve instructional delivery. According to Wetzell (1929), any measure of student outcomes to ascertain the effectiveness of education at this time, should be predicated on the integration of aptitude tests to determine ability level of a learner, development of specific objectives for the course, and reliable measures of learning.

**Post WWII focus on teacher evaluation.** Another movement regarding teacher evaluation emerged in the post-World War II era, according to Marzano et al. (2011). This movement hinged on the notion that evaluation should focus on the teacher as a unique individual. Whitehead (1952) provided six categories that teachers perceived as relevant, and claimed that administrators should pay more attention to the chief aim of education-effective teaching. Adding to this line of thought, Goldhammer (1969) established the clinical
supervision movement centered on the performance of the teacher in the delivery of a lesson. The focus of the five-phase process of analyzing one class session was for educators to collaborate in the development of their own evaluation. Cogan (1973) claimed that administrators should identify critical incidents that “impede desired learning in striking fashion” (p.172). Cogan (1973) elaborated that the teacher must be engaged in the supervisory process.

**Emphasis on the evaluator.** The 1980s saw the introduction of Madeline Hunter’s (1980) seven-step model of a lesson to be integrated with the clinical supervision model. Hunter promoted evaluation of lessons to include an anticipatory set, objectives and purpose, input from the teacher, modeling demonstrated by the instructor in conjunction with assessing the students for understanding, and providing guided and independent practice. Fehr (2001) found that many states adopted Hunter’s model of instruction. During the 1980s the RAND group (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984), conducted a study of 32 school districts that surmised that evaluation systems characterized as educational, developmental, or reflective lacked the specificity to truly improve pedagogical development. The narrative components of evaluation were found to result in a lack of uniformity in assessing teachers and schools. The RAND group cited identifiable problems with teacher evaluation. Primary among the problems were lack of competence of administrators to effectively evaluate sound teaching practice, teacher resistance to feedback, absence of uniformity and direction in the process of evaluation, and lack of training for evaluators.

**Integrating teacher evaluation and student growth.** In 1996, Charlotte Danielson introduced *Enhancing Professional Practice: A Framework for Teaching*, which was
subsequently revised in 2007. This work was described by Marzano et al. (2014) as the “reference point for any new proposals regarding supervision and evaluation given its past and current popularity” (p. 8). According to Marzano et al., Danielson attempted to capture the full-complexity and dynamic nature of classroom teaching. Danielson (1996) sought to honor the complexity of teaching, to provide a framework for the language of professional conversation, and to provide structure for self-assessment and reflection on the process of teaching. Marzano et al., contended that Danielson provided the most sophisticated model of evaluation to that time.

In 2011, the University of Chicago released a report that provided research-based evidence showing that new teacher observation tools, when accompanied by thoughtful evaluation systems and professional development, can effectively measure teacher effectiveness and provide teachers with feedback on the factors that matter for improving student learning. This was especially relevant for school districts implementing Charlotte Danielson’s (1996) framework for teaching. Pisciotta (2014) claimed that to improve student outcomes, teachers should encounter a system that serves as a professional development plan for improvement and simultaneously serves as a means of identifying teacher value. Lawmakers have legislatively empowered districts with the tools necessary to retain only the highest skilled educators by enacting legislation that arms districts with powers of accountability, yet they have failed to provide a blueprint of how to arrive at evaluating educators.

There is little debate that it is necessary to incorporate a viable means of integrating student achievement data into the determination of the effectiveness of an instructor. However, the literature has highlighted the absence of multidimensional educator evaluation
systems (Brandt, Mathers, Oliva, Brown-Sims, & Hess (2007); Learning Point Associates, 2010; Weisberg, Sexton, Mulhern & Kellog 2009).

**Teacher Evaluation in Michigan, 2011–Present**

Hu (2015) claimed that, as of 2010, 97% of educators in Michigan were rated as *Proficient* or *Highly Proficient* through local evaluation systems. The call for more accountability in respect to Michigan’s approach to public education was instituted by legislators with the implementation of amending the section of the Revised School Code relative to teacher evaluation. In 2011–2012, Michigan school districts began the development of educator evaluation systems to meet the requirements of important new legislation, Public Act 102 (PA, 2011)) passed by the Michigan State Legislature. This legislation aimed to evaluate all educators, thereby providing all students a high-quality educational experience and readiness for college and careers.

The call to measure effectiveness utilizing quantifiable student achievement data resoundingly increased across the State of Michigan, as student proficiency and growth models were scheduled to represent a minimum of 50% of a teacher’s evaluation score by 2015/16, according to PA 102 (2011). The student growth component is critical to a district complying with the intent of PA 102 (2011), and the connection to teacher impact upon student achievement is essential according to Lafee (2014). The evidence that teachers have a strong impact on student learning is in the research of Borman and Kimball, 2005; Hanushek, et. al., 2005; Nye, Konstantopoulos, and Hedges, 2004; and Muijs and Reynolds, 2002.

Michigan school districts have been mandated to comply with the PA 102 statute since 2011. The impact on school culture and the perceived efficacy of instructors relative to
this process has yet to be thoroughly explored beyond the case study of Aramath (2014), who researched two districts in Michigan that claimed to have incorporated evidence of student learning as a significant element of teacher performance. According to Aramath, the Education Alliance of Michigan (2011) and the Michigan Association of School Personnel and Administrators (2010) reported that relatively little research was available or existed that provided guidance to school districts in selecting and sustaining a viable educator evaluation model. Joshua Traughber (2015) added to the knowledge base in his doctoral dissertation at Eastern Michigan University entitled *An Exploration of How Middle School Principals Make Sense of Current Teacher Evaluation Policy and How Their Understanding Impacts Their Work to Help Teachers Develop Instructional Practice*.

**The Governor’s Council for Educator Effectiveness.** Public Act No. 102 of 2011 created the Governor’s Council on Educator Effectiveness, later changed to Michigan Council for Educator Effectiveness as a 2-year temporary agency, staffed and supported by the Governor's office to provide tools that improve teacher effectiveness. The council, which included five voting members, three of whom were appointed by Governor Rick Snyder, and one each by the Senate Majority Leader and Speaker of the House, was charged with submitting a report to the State Board of Education by April 30, 2012, that included the following objectives with supporting elements:

1. A student growth and assessment tool.

   - Is a value-added model that takes into account student achievement and assessment data, and is based on an assessment tool that has been determined to be reliable and valid for the purposes of measuring value-added data.
• In addition to measuring student growth in the core subject areas of mathematics, science, English language arts, and social science, will measure student growth in other subject areas.

• Complies with all current state and federal law for students with a disability.

• Has at least a pre- and post-test.

• Is able to be used for pupils of all achievement levels.


• In addition to the student growth and assessment tool, the recommended state evaluation tool for teachers may include, but is not limited to, instructional leadership abilities, teacher and pupil attendance, professional contributions, training, progress report achievement, school improvement plan progress, peer input, and pupil and parent feedback.

• The council shall ensure that the recommended state evaluation tool for teachers will allow all special education teachers to be rated.

• The council shall seek input from school districts, intermediate school districts, and public school academies that have already developed and implemented successful, effective performance evaluation systems.

3. A state evaluation tool for school administrators described in subsection (3).

• In addition to the student growth and assessment tool, the recommended state evaluation tool for these school administrators may include, but is not limited to, teacher and pupil attendance, graduation rates, professional contributions, training, progress report achievement, school improvement plan progress, peer input, and pupil and parent feedback.
4. Recommended parameters for the effectiveness rating categories for teachers) and for school administrators.

5. Recommended changes to be made in the requirements for a professional education teaching certificate that will ensure that a teacher is not required to complete additional postsecondary credit hours beyond the credit hours required for a provisional teaching certificate.

6. A process for evaluating and approving local evaluation tools for teachers and school administrators.

The legislature planned to review the report submitted by the governor's council on educator effectiveness and to enact appropriate legislation to put into place a statewide performance evaluation system taking into consideration the recommendations contained in the report. However, the timeline for the council to deliver the comprehensive report was delayed until July of 2013, as the council found the task to be lengthy and cumbersome.

The Center for Educational Performance and Information (CEPI) was established by the Michigan Legislature (Sect. 94a of the State School Aid Act) to coordinate the collection, management and reporting of education data from early childhood through K-12 and postsecondary education and into the workforce. CEPI facilitates efficient data gathering to reduce the administrative burden on reporting entities while ensuring student privacy.

In the first two years of educator evaluation implementation after the new legislation, school districts were allowed to redesign, revise, or maintain their existing systems, as long as the systems met basic requirements. The variation in systems made direct comparison of district effectiveness ratings and systems extremely difficult, because the criteria used to determine effectiveness in District A could be very different than in District B.
2011–13 Survey–Educator Evaluations and Effectiveness in Michigan

Responding to requirements of PA102 (2011), the Council for Educator Effectiveness established the Michigan Department of Education’s (MDE) 2011–13 survey of local districts entitled Educator Evaluations and Effectiveness in Michigan. Seven hundred seventy school districts provided information in five sections about how they approached K-12 evaluation systems. The first four sections focused on components used by districts to evaluate educators. The last section involved how district administrators used data to guide year-end evaluation decisions.

Findings of the survey indicated that considerable variation existed among districts in respect to the various observation tools utilized for evaluation and the weighted measures for each category. Overall, there seemed to be little relationship between the state labels of priority or reward and the effectiveness of a school or of teachers as defined by the evaluation instrument. Teachers rated as ineffective were just as likely to be employed at a Rewards School as they were a Priority School. A Rewards School would be rated by the Michigan Department of Education (MDE) as being in the top 5th percentile regarding pace of improvement or achievement in comparison to districts across the state. A Priority School is in the bottom 5th percentile when compared to all the districts in the state.

In 2013, only 69.1% of schools reported that for the two years that survey results were collected they used student growth data in year-end evaluations as required by PA 102 (2011). Just more than half of K-8 schools reported using state test achievement results. The most common categories for evaluation were instructional practice, classroom management practices, pedagogical knowledge, and content knowledge. The survey data also indicated that female teachers with master’s degrees or higher and language arts teachers were most
likely to be rated highly effective. Minority teachers were more likely to be rated highly effective than White teachers.

**Four Observation Tools Piloted**

Led by the University of Michigan, Institute for Social Research in 2013, the Governor’s Council for Educator Effectiveness initiated a pilot of four observation tools in a program entitled *Promoting High Quality Teacher Evaluations in Michigan.* Thirteen school districts selected to participate included Big Rapids, Cassopolis, Clare, Farmington, Garden City, Gibraltar, Harper Creek, Leslie, Marshall, Montrose, Mt. Morris, North Branch, and Port Huron.

The four evaluation tools reviewed included Danielson’s framework for teaching, 5 Dimensions of Teaching and Learning, Marzano’s teacher evaluation model and the Thoughtful Classroom framework. Although the domains outlined by Danielson (1996) seemed to be pivotal to the skills teachers must exhibit to demonstrate good instruction, legislators directed educators to consider student learning as the true objective.

**Danielson framework.** Danielson (2011) was mindful that the increased relevance of accountability provides a much different climate than the initial framework that she provided in 1996. Danielson created her tool as a means for evaluation to serve as a way of professional development. Weschler and Shields’ (2008) research indicated that, aside from family background, the biggest determinant of learning is a child’s teacher. Considering the ambiguity inherent in the complexity of ascertaining teacher effectiveness and the connection to student learning, an in-depth look at Danielson’s framework for teaching is important, as it pertains, to student learning, quantifying teacher effectiveness, and defining success.
The review of the framework created by Danielson (1996) and its underpinnings appears to be critical to understanding the process of teaching and learning with the accountability that stakeholder’s mandate. Kane, Tyler, and Wooten (2011) found that looking solely at test measures offers little insight into improvement. Further, concentrating solely on standardized tests will evolve into teachers ignoring learning and focus on test-taking skills. The four domains highlighted by Danielson offer a connection to assessment and an impartial approach to evaluating teacher performance leading to effective teachers and good instructional practice. Although evaluation may be thought of as something that is “done to teachers,” Danielson (2001) advocated that the process should become collaborative and, in turn, serve as a means for teachers to provide evidence of effective instruction.

According to Danielson, quality control is the only area that lawmakers are concerned with in the evaluation process, but to truly improve educational opportunities for students, teachers need a blueprint from which to work. This notion was further advanced by the work of Opfer and Pedder (2010), who conducted research into the process of teacher-learning and the sub-systems involved in the process of teachers enhancing their skill sets. The domains of educator effectiveness outlined by Danielson (1996) serve this function: 1) planning and preparation, 2) classroom environment, 3) instruction, and 4) professional responsibility.

Six components compose the domain of planning and preparation:

- demonstrating knowledge of content and pedagogy,
- demonstrating knowledge of students,
- selecting instructional goals,
- demonstrating knowledge of resources,
- designing coherent instruction, and
• assessing student learning.

Danielson constructed the planning and preparation domain around the work of Shulman (1988), who proposed that the distinction between the knowledge base of the instructor and transferring the content to the form where it is presented in a pedagogically powerful manner to a level that learners can understand is essential to learning. Reynolds (1992) surmised that competent teachers create lessons that enable students to interact and connect with new ideas and concepts. A definition of learning would seem essential to understanding the nexus to teacher effectiveness in preparing students to learn. McCombs (1992) referred to learning as an individual process of constructing meaning from information and experience. This combination of information and experience is then transposed through each individual’s unique perceptions, thoughts, and feelings. Jones (1992) found that establishing clear learning goals is instrumental to delivering instruction that is coherent for learners so that students can effectively demonstrate understanding of new content.

Assessing learning is essential for an effective teacher. Cryan (1986) found that constructed responses, interviews, and performance assessments are more accurate measures of understanding and comprehension than a single standardized approach. The classroom environment is the second Danielson (1996) domain.

Five components compose the domain of classroom environment:

• creating an environment of respect and rapport,
• establishing a culture of learning,
• managing classroom procedures,
• managing student behaviors, and
• organizing physical space.

Makela (2010) conducted a study of students in a Colorado history class. She was able to demonstrate a high impact on environments where students were able to integrate group learning. The most effective characteristics present in the classes were personal interaction, focus on social and collaborative skills, interdependence on finding solutions, monitoring of progress by the instructors and the students, and personal accountability for stakeholders. These findings mirror the underpinnings of Danielson’s establishment of the connection between learning and the ambiance in which students engage in the educational practice. Group and team learning is often reflected in teacher issued grades.

Danielson (1996) explained that first-year teachers must concentrate on environment above all. This was validated 13 years later by Boyd, Grossman, Lankford, Loeb, and Wyckoff (2009) in their study of New York City teachers. They found that student achievement can be altered based on teacher training in the realm of environment in the first year. Studies on specific rates of effectiveness for time-on-task, student participation, engagement, and appropriate behavior have been conducted by Good (1984). Doyle (1986) was also able to establish the need for effective teachers in establishing routines and procedures, even as a precedent to the academic content expectations. A positive aura of morale, expectations, and enthusiasm also has a significant impact on student learning according to Tornatzky and Pettigrew (1974). The aforementioned research paved the landscape for the pillars of Danielson’s (1996) work of building the rubric to reflect the connection of affective relationships that must be met before cognitive learning can be cultivated. Instruction is Danielson’s third domain.

Five components compose the domain of instruction:
- communicating clearly and accurately,
- using questioning and discussion techniques,
- engaging students in learning,
- providing feedback for students, and
- demonstrating flexibility and responsiveness.

Teaching and learning continues to evolve as evidenced by the work of Kinnebrew, Segedy, and Biswas (2014), who found that the metacognitive learning of students can be influenced and altered by the integration of technology environments in which students find themselves. No Child Left Behind (NCLB, 2001) legislation emphasized the individual needs of students. Instruction is moving toward differentiation to each student, as indicated by the research of Dixon, Yssel, McConnell, and Hardin (2014), who found that teacher efficacy in student results proves critical to differentiated instruction. It should be noted that individualized conceptual learning provides the origin of instruction as detailed by Danielson (1996).

Ellet (1990) signified that learner participation, involved with discussion and engagement, is needed if effective learning is going to occur. The communication of goals and the consideration of the students in establishing the goals were pivotal to the study of Edmonds and Frederickson (1978). In the arena of mathematics, Rosenshine and Furst (1971) were able to establish a positive relationship between teachers who gave frequent and accurate feedback in relation to student achievement. Teachers are as much managers and facilitators as deliverers of content. According to the *Educator Evaluations and Effectiveness in Michigan* (2013) survey of 707 district respondents, results released by MDE found that 488 districts utilized the Danielson (1996) framework tools used in local

Six components compose the domain of professional responsibility:

• reflection on teaching,
• maintaining accurate records,
• communicating with families,
• contributing to the school and district,
• growing and developing professionally, and
• showing professionalism.

The National Board of Professional Teaching Standards (1991) identified three components that are consistent with advanced teachers: (a) teachers are committed to students and learning, (b) teachers reflect on their practice, and (c) teachers are members of learning communities. The idea that teachers are professionals has not been in existence for as long as one may think. However, Calderhead (1989) found that teachers who reflect on their professional practice are considerably more effective than teachers who do not view what they are doing in that same contextual lens. This is a critical piece—reflection within the responsibility to the profession, as highlighted in the Danielson (1996) rubric.

Collaboration is also an area that is essential to the idea that mutual effort will produce student learning outcomes more effectively. The collaboration is not exclusive to professional colleagues sharing ideas, but also inviting community members, parents and students to get involved in the learning process. Griffin (1986) conveyed that an effective teacher interacts with students, colleagues, and community members purposefully and effectively. Griffin (1986) relayed that effective teachers view teaching as more than
meeting with students, and they work with colleagues to identify and act on problems in the classrooms and schools.

**Post-Danielson era of evaluation.** With the introduction of Tucker and Stronge’s 2005 book, *Linking Teacher Evaluation and Student Learning*, came a strong advocate for the inclusion of student achievement data to be integrated into the evaluation process. Tucker and Stronge contended that it is the data harvested from the students that should dictate and drive the assessment and instructional competency of educators. Their work parlayed into Toch and Rothman’s (2008) report, *Rush to Judgement*, and Weisberg, Sexton, Mulhern, and Keeling’s (2009) study titled *The Widget Effect*. Both studies found that the evaluative processes in the country as a whole were not designed to promote educational improvement. Only 14 states mandated that schools annually evaluate teachers, and some evaluative processes were nothing more than *satisfactory or unsatisfactory* in their format, according to Toch and Rothman (2008). Weisberg et al. (2009) found that “73% of teachers surveyed claimed that their most recent evaluation failed to identify any development areas” (p. 6).

**5 Dimensions of Teaching and Learning.** The Center for Educational Leadership (CEL, 2007) developed the 5 Dimensions of Teaching and Learning from multiyear research efforts at the University of Washington. The instructional framework defines the core elements of effective teaching and learning along five dimensions: purpose, student engagement, curriculum and pedagogy, assessment student learning, and classroom environment and culture. Website information highlights 30 indicators of teacher performance and an online assessment tool that “measures leaders’ ability to observe and analyze instruction, offers timely feedback to teachers, and guides teachers’ learning.” The
report of *Educator Evaluations and Effectiveness in Michigan* (2013) showed that 35 of the 707 reporting districts in the State of Michigan chose the 5 Dimensions of Teaching and Learning as the focus of their teacher evaluation plan.

**Marzano’s causal teacher evaluation model.** The Michigan Department of Education: *Educator Evaluations and Effectiveness in Michigan* (2013) reported that 176 districts chose Marzano’s tool for teacher evaluation, second behind Danielson’s 488 districts. The Marzano’s causal teacher evaluation model was developed in 2012 for the Washington State Superintendent of Public Instruction and the Washington Teacher Principal Evaluation Pilot. Marzano’s most recent model follows from a synthesis of the research and theory in previous work including the following: *Classroom Instruction That Works* (Marzano, Pickering, & Pollock, 2001), *What Works in Schools* (Marzano, 2003), *Classroom Assessment and Grading That Works* (Marzano, 2006), *The Art and Science of Teaching* (Marzano, 2007), and *Effective Supervision: Supporting the Art and Science of Teaching* (Marzano, Frontier, & Livingston, 2011). Marzano’s causal teacher evaluation model encompasses four domains with 60 sub sets. The four domains are: (a) classroom strategies and behaviors, (b) preparing and planning, (c) reflecting on teaching, and (d) collegiality and professionalism.

**The Thoughtful Classroom.** Developed by Silver Strong & Associates in 2007, the Thoughtful Classroom model comprises three domains with ten related dimensions. The first domain, four cornerstones of teaching, emphasizes the organization of the work of teaching and learning and the culture of the classroom stemming from positive relationships and sharing the learning experience. The second domain, effective instruction, stresses the structure of the learning process—preparing for new learning, presenting, reinforcing and
applying, and reflecting on new knowledge. Finally, the third domain, looking beyond the classroom, focuses attention on effective professional practice and the role of professional development in improving student achievement. The Michigan Department of Education’s Educator Evaluations and Effectiveness in Michigan (2013), reported that only seven districts in the State of Michigan invested in The Thoughtful Classroom teacher evaluation tool.

**Student Growth Tools Piloted**

While Danielson (1996) and others made significant strides in teacher evaluation models the federal initiative, No Child Left Behind Act (NCLB, 2001)), aimed at increasing academic achievement as the focus of school improvement and required that all students and subgroups of students be assessed annually. In Michigan, four student growth tools were piloted: (a) Northwest Evaluation Association (NWEA) for Grades K-6 (b), ACT Explore for Grades 7-8, (c) ACT Plan for Grades 9-10, and (d) ACT for Grade 12. Porter’s (2002) research determined that pupils make more academic gains when instruction is effectively connected to assessment. The challenge appeared to be to determine what system-wide student assessment best correlated with teacher evaluation methods.

Survey results 707 district respondents released from the office of Educator Evaluations and Effectiveness in Michigan (2013) showed that 33% of school districts used the NWEA assessment tools. This was an increase of 10% from 2012. Northwest Evaluation Assessment [NWEA] is an organization that created computer-based assessments that identify projected rates of growth for students in comparison to students of similar academic ability across the nation. NWEA gathers information from more than 4,400 districts across the country to develop baseline information. Students take computerized adaptive tests with the difficulty adjusted based on the performance of the student. Students
can be tested up to four times a year. Testing from fall to spring and from year to year provides a baseline for instructors to determine the students’ level when they come to a teacher and their level when they leave the teacher. Scores reported include percentile scores, achievement scores, and growth scores. Students in grades K-12 are eligible to be tested in the areas of math, language usage, reading, and science. The value and utility of incorporating NWEA assessment is that data account for student growth, and further, NWEA provides a comparison against the nation. As stakeholders put increased emphasis on competing globally, the national averages should be considered.

**Key Findings of Pilots of Teacher Evaluation and Student Growth Tools**

The key findings of the piloting process found that most districts lacked policy to address PA 102 (2011) and that districts worked as teams to integrate appropriate policies. Principals in all pilot districts acted as the evaluators, but teachers played a critical role in producing student achievement data. Principals almost unanimously reported that they conducted pre- and post-observation meetings, and 82% of teachers reported participating in these meetings. Principals reportedly averaged 248 hours per year or the equivalent of 31 full work days conducting evaluations. Further, only 60% of principals felt confident conducting observations with the vendor-provided tool, citing minimal training in respect to implementing the tool.

In respect to student growth, findings showed how difficult it would be to incorporate state assessment data from the Michigan Educational Assessment Program (MEAP) into teacher evaluations. Only one-third of Michigan teachers entered data into the MEAP. The reporting of scores and timing of the year for testing also proved to be a heavy burden. The most common measures of student growth in pilot schools were locally created tests devised
by teachers (54% of elementary schools, 69% of middle schools and 64% of high schools used common assessments).

The piloting process revealed that the most common approaches to determining student growth and achievement were not highly technical or sophisticated systems from the point of a psychometric analysis. The student growth tools provided externally to schools were largely ignored and not implemented. Only 20% of elementary teachers reported that they included NWEA data in their evaluations. Multiple assessment tools created a canvass of non-uniform measures to analyze student growth. The pilot process also unveiled that varying cut scores and weights assigned to different categories of evaluation and growth measures made it very challenging to compare district ratings.

Findings of primary importance from the piloting process showed that 89% of principals believed that the evaluation tools they tested focused on integral aspects of teaching, whereas only 62% of teachers believed that ratings assigned to them were accurate, and only 47% of teachers reported the tools easy to understand.

**Recommendation from Governor’s Council**

Upon the completion of pilot tests of teacher evaluation models and student assessment methods, the final recommendations of the council’s executive summary (2013) charged the state with the following:

- Evaluate the quality of third-party or locally developed assessments.
- Train teachers to apply adequate rigor in the development and measurement of student learning objectives.
- Produce value-added modeling (VAM) scores for educators on state-provided assessments in the core content areas in all cases that are possible and
professionally responsible.

- Base at least half of the student growth component on state-provided VAM scores for teachers in core content areas in grades for which there is growth data available from state-mandated assessments (currently reading and mathematics in grades 4–8, but likely to change over time).

- Continue to develop or select, as well as support, all aspects of assessments that are aligned to state-adopted content standards in all core content areas (i.e., English language arts, mathematics, science, and social studies). The state should also develop or select assessments aligned to state-adopted content standards in high-volume non-core content areas where state-adopted content standards exist (e.g., arts, health and physical education, career and technical education, and many high school electives).

- Acknowledge that adoption of state-provided VAM or growth data for teachers of other subject areas should be at the discretion of local education agencies [LEAs].

- Consider the use of state-provided or growth data in core content areas may be used in a teacher’s evaluation using information from that teacher’s students, even if the teacher does not teach in one of the core content areas. This means that teachers may be evaluated, in part, for the learning of their own students, even in subject areas that they do not directly teach. This may be done as long as the teacher knows that he or she is expected to be contributing to students’ growth in core content areas, and there is a reasonable connection of the core content to the teacher’s teaching assignment.

- Promote collective work on instruction within school buildings, school-level by
use of VAMs for individual teachers’ evaluations if there is a reasonable connection of the core content to the teacher’s actual teaching assignment. This means that teachers may be evaluated, in part, for the learning of students whom they do not directly teach. However, school-level VAMs may not comprise more than 10% of the individual teacher’s student growth component.

- Evaluate teachers based on alternate measures of student growth that meet the guidelines for rigorous and appropriate assessment of student growth in the applicable subject area in content area assignments for which there is no state-provided VAM or growth data available.

The governor’s council failed to address the data component which is slated to constitute 50% of the overall evaluation score as of the 2015-16 school year. In essence, the state remained silent on the most significant issue in respect to identifying a means to accomplish their suggestions and largely allowing this element of the law to be the responsibility of local districts.

However, numerous bills have been introduced in the Michigan Legislature that address the specific percentages of student achievement data that constitutes a teacher’s evaluation. Two bills that were introduced were not enacted in 2014 (Feldscher, 2015). Senator Phil Pavlov, R-St. Clair, introduced Senate Bill 103 enacted on October 21, 2015 that reduced the percentage of student data to 25% for 2017-18 and moves to 40% in 2018-19. Pavlov’s bill affords local educational entities the opportunity to create their own evaluation tools. In committee hearings, a concern was raised in the testimony of Dr. Paul Salah of Wayne RESA that in 2018-19 the data component would be comprise 20% state assessments and 20% local assessments (Feldschler, 2015). The issue, according to Dr. Salah, is that the
state has yet to create and implement measured, valid, and reliable assessments.

**Change and Innovation**

Attempts of Michigan school districts to determine a process of teacher evaluation integrated with student achievement assessment in compliance with PA 102 (2011) have introduced an era of change and innovation throughout the state. Rogers’s (1983) change model offered an explanation for the plethora of different roles that could surface among individuals in an organization in the process of change. According to Rogers,

> One kind of uncertainty is generated by an innovation, defined as an idea, practice, or object that is perceived as new by an individual or another unit of adoption. An innovation presents an individual or an organization with a new alternative or alternatives for solving problems. But the probabilities of the new alternatives being superior to previous practice are not exactly known by the individual problem solvers. Thus, they are motivated to seek further information about the innovation in order to cope with the uncertainty that it creates. (p.17)

Due to the uncertainty in a period of innovation and change, there is an important need for the dissemination of information to individuals in an organization. Rogers (1983) calls this phenomenon diffusion. Rogers found that diffusion is the process by which an innovation is communicated over time through certain channels among the members of a social system. It is a special type of communication that is relative to new ideas.

Communication is a process in which participants create and share information with one another in order to reach a mutual understanding, rather than when a message is simply transferred from one to another. Individuals adopt change at different rates. Rogers (1983)
described four characteristics of adoption with which individuals identify when presented with innovation:

1. Relative advantage is the degree to which an innovation is perceived as better than the idea it supersedes. The degree of relative advantage may be measured in economic terms, but social-prestige factors, convenience, and satisfaction are also often important components. It does not matter so much whether an innovation has a great deal of “objective” advantage. What does matter is whether an individual perceives the innovation as advantageous. The greater the perceived relative advantage of an innovation, the more rapid its rate of adoption is going to be.

2. Compatibility is the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters. An idea that is not compatible with the prevalent values and norms of a social system will not be adopted as rapidly as an innovation that is compatible. The adoption of an incompatible innovation often requires the prior adoption of a new value system. An example of an incompatible innovation is the use of contraception in countries where religious beliefs discourage use of birth-control techniques, as in Muslim and Catholic nations.

3. Complexity is the degree to which an innovation is perceived as difficult to understand and use. Some innovations are readily understood by most members of a social system; others are more complicated and will be adopted more slowly. For example, the villagers in Los Molinos did not understand germ theory, which the health worker tried to explain to them as a reason for boiling their drinking water. In general, new ideas that are simpler to understand will be adopted more rapidly than innovations that require the adopter to develop new skills and understandings.
4. Trialability is the degree to which an innovation may be experimented with on a limited basis. New ideas that can be tried on the installment plan will generally be adopted more quickly than innovations that are not divisible. Ryan and Gross (1943) found that every one of their Iowa farmer respondents adopted hybrid-seed corn by first trying it on a partial basis. If the new seed could not have been sampled experimentally, its rate of adoption would have been much slower. An innovation that is trialable represents less uncertainty to Diffusion of Innovations Elements of Diffusion the individual who is considering it for adoption, as it is possible to learn by doing (p. 15).

Rogers and Shoemaker (1971) identify five potential roles individuals may be characterized into regarding change: (a) innovator: eager to try new ideas, open to change, willing to take risks, usually perceived as naïve or a little crazy and therefore not well integrated into the social structures; (b) leader: open to change, but more thoughtful about getting involved; trusted by other staff, and sought for advice and opinions; (c) early majority: cautious and deliberate about deciding to adopt an innovation and tends to be a follower, not a leader; (d) late majority: skeptical of adopting new ideas and “set in their ways;” can be won over by a combination of peer pressure and administrative expectations; (e) resister/laggard: suspicious and generally opposed to new ideas, usually low in influence and often isolated from the mainstream. The basis upon which this framework is devised hinges on the notion that there are very few innovators and a small number of leaders and resisters, but the majority of participants in an organization could be described as early majority or late majority.
Teacher Evaluation in the District in this Study

The current political climate surrounding the issues of teacher evaluation and assessment of student achievement accelerated the need to ascertain what stakeholders in the selected school district view as reliable data to determine the basis for teacher evaluation and acceptable rates of learning. The research activities in this study included interviews with the former superintendent, two board members, three principals, and ten teacher-leaders. Further data were gathered in evaluation policy documents, and other archived artifacts relative to the teacher evaluation system in the district.

The Danielson framework (1996) was adopted by the district in this study to evaluate teacher performance. According to Rowan et al. (2013), the Danielson’s rubric is the most widely utilized instrument by districts in Michigan.

Learning outcomes were scrutinized to assess if the InFORMED rubric designed by the district was an appropriate mechanism to define success in the realm of student learning outcomes, as identified by Danielson (1996), and whether stakeholders in the district interpreted the process of teacher evaluation to have played a role in increased achievement as identified by the Michigan Department of Education (MDE, 2013).

According to the Educator Evaluations and Effectiveness in Michigan (2013) survey of district respondents, results released by MDE found that fewer schools were using state assessment data in 2013 than they were in 2012 (72.9% in 2012 to 61.9% in 2013). The assumption would be that for 2014-15 even fewer districts, if any, integrated state testing. The State of Michigan shifted from MEAP to M-Step and intends to defer to SAT as opposed to ACT for 2015-16. The lack of familiarity with the results these assessments will yield
would leave districts shooting in the dark in respect to assuming what the data would look like.

Currently, the district being researched no longer includes state assessment data for individualized instructors for the MEAP/M-Step and MME. This is because confusion at the state and bureaucratic levels have resulted in a lack of clarity on what testing tool to utilize, when to test, and, how results will be shared/tracked.

Summary

This literature review offered an overview of issues and perspectives related to the integration of teacher evaluation and assessment of student achievement. A history of ideas and activities emanating from the federal government and the State of Michigan in particular set the stage for the most recent era of state legislation and the mandates of the Governor’s Council for Educational Effectiveness. Four teacher evaluation models and methods of assessing student achievement were piloted and recommendations following provided guidance to school districts as they attempted to comply with the objectives for teacher evaluation. The work of Everett Rogers (1983) was incorporated into the review of literature to describe a pattern of the adoption of change likely to occur in school districts in Michigan. A brief review of issues and strategies in the implementation of the state-mandated teacher evaluation process in the school district in this study offer an introduction to the specific methods employed in the conduct of this case study.
Chapter Three: Research Design and Methods

This study examined how one school district in northern Michigan responded to changing expectations regarding teacher evaluation and how the district revised their evaluation process to conform to one of the state mandated models, and simultaneously quantify student achievement data to account for 25% of a teacher’s evaluation.

Research Questions

The study was guided by the following research questions:

1. How did one school district prepare for the implementation of Public Act 102?
2. What strategies were employed to construct and sustain support for the modified teacher evaluation?
3. What did the district learn from the modified Danielson rubric and the InFORMED data results rubric?

Research Tradition

This qualitative, single case research was based on interpretivist epistemology. Gall et al. (2005) explained that social reality is viewed as a set of meanings that are developed by individuals who are actively engaged in the reality where the phenomenon is transpiring. Noting how legislation affected stakeholders in the district would be an example of this research approach. Study participants developed meaning as they reacted to change in the evaluation process in their school district brought about by state legislation PA 102 (2011).

The main instruments integrated for investigation were heavily saturated in field-based inquiry. Questions asked were relative to communication from the organization and individual’s responses in regards to the development and implementation of the change process using the basic format of the concerns-based adoption model (CBAM) developed by
the American Institute for Research (1978). This will be done in an effort for respondents to identify their stages of concern as the new process of evaluation was rolled out in respect to communication. Respondents spoke to the levels of use once the instrument was integrated as the district’s evaluation system.

The researcher utilized this case study to describe, explain, and evaluate the implications of PA 102 (2011) on a district. The application of the case study yielded insight into how the state should consider legislative impacts that have been instituted with little direction and no funding from the State of Michigan. Further, this study can lend perspective for fellow educators on how to grapple with the task of integrating meaningful teacher evaluation practices. Identifying what paradigms exist relative to teacher evaluation unveiled themes that germinated from the process of integrating a teacher evaluation instrument within a district.

The research attempted to reflect on how the evolution of teacher evaluation in Michigan since 2010 has been an experience that warrants investigation into a study of the phenomenon in an in-depth manner, in a specific instance, within its natural context and that accurately represents the perspective of the researcher and the participants as suggested by Gall et al. (2005). By looking at the particular illustration of how the district was compelled to alter its former practice of evaluation and move to reflect the intent of PA 102 (2011), the study focused on how key stakeholders responded to the challenges inherent in the mandate. There was a considerable amount of data to collect and analyze as there is a tremendous volume of evaluation data, artifacts, laws and verbal statements to obtain as this was an in-depth study designed to address an inclusive record of the response of one northern Michigan school district’s approach to evolving expectations implied by PA 102 (2011). Kirk and
Miller (1986) lend great support to the notion that researchers interact with research participants in their natural environment as this study intends to do. That said, the researcher was involved in the creation of the InFORMED data results rubric and the evolution of the tool for its complete existence. As such, the researcher was conscious to delineate between his own emic perspective and that as an outside investigator with etic perspective. As Bodgan and Bilken (2003) argue, the researcher exerts a great deal of time and effort acquiring and reviewing data that the researcher must always be cognizant of their own opinions. It is not likely that all biases can be removed relative to the interpretation of the data. Creswell (2009) would suggest that by considering the researcher’s own experience it can translate into more depth transpiring in the study.

The selection of this case was strategically identified as the district research embodies an extreme case of immersion as defined by Miles and Huberman (1994). The district earned a reputation for exerting considerable effort to meet the intent of PA 102 (2011). The district was a forerunner of incorporating student achievement data into the evaluation process and was utilized as a model by many other districts. The Education Trust-Midwest (2012) recognized the district along with the Detroit Free Press in November of 2012 for the district’s response to PA 102 (2011).

Case Study Research

Yin (2003) claims that case study research is most useful and a preferred strategy to investigate real life context situations when the researcher is committed to answer the “how and why” questions of the implications that phenomenon have had on circumstances. Yin (2003) further contends that there are two types of case studies that can be conducted: (a) exploratory and (b) descriptive. For this exercise, this researcher intends to perform a
descriptive analysis of the phenomenon of the impacts that legislation relative to teacher
evaluation has had on one Northern Michigan School District. Yin (2003) feels that the
strength of a case study is inherent in the ability to contend with a wide array of evidence
such as documents, artifacts, interviews, and observations beyond what would exist in a
historical study. Schramm (1971) defined a case study as appropriate to utilize when trying
to illuminate a decision or set of decisions: why they were taken, how they were
implemented, and what the result was.

Stocker (1991) identified a case study as a research strategy as an all-encompassing
method inclusive of the logic of design, data collection techniques, and approaches to data
analysis. Yin (2003) highlighted that case study is preferred when looking at contemporary
events. The impact and current implications on the field of education relative to educator
evaluation and Public Act 102 facilitate the call for a case study of a rural Michigan district.
Yin (2009) contended that the case study attempts to examine a phenomenon in its real life,
especially when the boundaries and context may be ambiguous. Attention, therefore, is
devoted to and “described within the perspective of the context surrounding it” with the
aspiration that a deeper and more intimate understanding can be cultivated (Krathwohl, 2009,
p. 353). This study serves as a singular case study as defined by Yin (2009).

Setting for the Study

In an effort to provide contextual completeness as defined by Yin (2003), the
researcher worked to embrace a comprehensive analysis of the district. The system of study
that was researched finds its setting in a district that is located in northeastern lower
Michigan in a rural community that is approximately 200 miles north of Detroit and situated
along the coast of Lake Huron. The district comprises two buildings. Administrative office
as well as middle school and high school students attend one building serving 7th–12th grade. Preschool through sixth grade are housed in another building. The district has four administrators, 72 teachers, and 64 and one half other employee positions.

The district’s self-analysis in the (2014–2015) School Improvement Plan claimed to be “an exciting place in respect to its ability to offer quality programs” to students despite the economic and social constraints that confront the community. In 2014–2015, the district claimed a 78% free and reduced lunch rate, as reported by the State of Michigan. The district has had rapid growth as evidenced by MI School Data, reporting that the district attained an increase of 18.99% in its graduation rate from 2011 to 2013. The district’s high school was recognized as a Rewards School by the State of Michigan for two consecutive years, 2014 and 2015. Being designated a Reward’s School meant that the Michigan Department of Education identified a school as experiencing more academic growth than at least 95% of the schools in the state. The elementary school in the district received a Green Rating, the highest growth score possible, post 2011 when the Top–Bottom Rankings released by MDE had the district’s high school in the 4th percentile; this was in contrast to 2014 when the district ranked in the 53rd percentile. The district was plagued by an apparent disconnect and lack of trust among teachers and administration prior to 2011. Drastic measures were instituted in the school district to achieve the amount of movement academically attained, including an emphasis on the data scores of students, and the relationship to an individual teacher’s evaluations.

Altheide and Johnson (1994) encouraged the case study researcher to incorporate multivocality as participants in the study possess unique and diverse perspectives and insight with tacit knowledge that the individuals may not have the capability to express with words.
Incorporating a diverse group of stakeholders impacted by the institution of the teacher evaluation the researcher is striving to be complete and credible. The duration of time studied extends from the spring of 2010 to the conclusion of the 2014-15 school year. There were voluminous data to review of the long-term observation period. As recognized by Gall et al. (2005), the attitudes, feelings, and perspective of individuals in 2017 may vary from the inception of Public Act 102 in 2010 after being exposed to the phenomenon over time. The researcher integrated numerous interviews and perspectives of stakeholders, including staff who retired as well as those who were laid off due to their rankings on the evaluation instrument. This representativeness check served to mitigate the potential of elite informants providing a disproportionate amount of data to the study.

**Research Methods**

The researcher examined how one northern Michigan district addressed expectations for teacher evaluations. Notes and video/audio recordings served as data to analyze. Participants, including a former superintendent, two board members, three principals and ten teacher-leaders were invited to participate in open-ended questions in an interview format (see Appendix A). Documents and media analysis were reviewed with the objective of garnering an accurate assessment of the evolution of the evaluation phenomenon in the district and the impact and influence the legislation has had. The researcher collected data relative to the modified Danielson rubric and InFORMED from 2011 to the end of the 2014–2015 school year through STAGES. STAGES exists as a web-based system designed to automate staff evaluation, and is a division of K-12 Evaluation Solutions®. STAGES was designed by Saginaw Valley State University in 1999 by means of a U.S. Federal Teacher Quality Grant. STAGES has since been utilized by districts across the United States as a
database to amass data utilized to evaluate teachers, administrators and support staff. The researcher collected and analyzed student achievement data through InFORMED data results. This software integrates student growth and achievement data for an individual educator’s evaluation. InFORMED was designed by the district and currently incorporates local and national achievement results. This tool was designed to be a customizable web-based instrument which allows users to choose which data pieces to integrate into the evaluation process. A review of artifacts such as planning documents, board updates, media reports and personal documents of key stakeholders was included in representing how the district worked to create InFORMED and the modified Danielson rubric. A review of self-report measures such as self-evaluation tools and reflective data from principals and teachers was immersed into the research.

Identifying key educational stakeholders and their role conforming to PA 102 (2011) lent great insight into the impact of this legislation and the far-reaching implications of the force upon which a district had to move to adhere to the law. Participants in the study were the former superintendent, technology director, two board members, three former principals and ten teacher-leaders. The former superintendent was included in the study because it was her vision and leadership that led to the construction of the InFORMED rubric. The technology director was incorporated, as her management and role in the evolutionary capability of the process was very significant. Three former principals were included in the study as each had significant input into the creation and evolution of the process of teacher evaluation in the district. They all worked in other areas of evaluation and could offer unique insight and perspective relative to a contextual vantage point. One former school board member and one current member were selected to give a historical balance to the study. Ten
teacher leaders were selected; five teachers have retired or recently moved to other districts. The other five teachers were members of the union’s professional negotiation team. This was done to mitigate the potential impact of bias that could enter into the research.

**Procedures and Collection of Data**

Data were gathered through interviews of stakeholders in the field and by reviewing archived artifacts relative to the teacher evaluation system integrated by the district. This was completed in conjunction with the review of pieces which serve as data to substantiate interviews. A review of the database system InFORMED also transpired for each of the teachers interviewed. Interviews were conducted in a face-to-face manner. Yin (2009) articulates that the interview component of the study is integral to the collection of data. Participants were asked focus questions as described by Yin (2009), but interviews were conducted in a manner that allows for open dialogue. There was a specific set of inquiry, but it allowed the participants to describe in their own words the phenomenon from their perspective. The protocol for interviews of teacher-leaders, administrators, school board members, and the technology director are shown in Appendices C through F. Notes were taken during the interviews, and the interviews were also digitally recorded.

Data were procured through accessing the teacher evaluation software utilized by the district, scheduling interviews with key stakeholders, and reviewing documents. Yin (2009) advised that a strong chain of evidence be established for a case study to be followed from the research question to the conclusion. There was an audit trail that is inclusive of the sources for recording data, process notes, and the development of instruments to procure data and any data reduction processes and data reconstruction products. The researcher strived to foster authenticity, as described by Alder (1994), by accurately portraying the challenges
critical educational stakeholders played in the construction of the evaluation tool and its implementation from 2010 to the end of the 2014–2015 school year.

By disconfirming case analysis, as suggested by Miles and Huberman (1994), an extreme case of a district that attempted to meet the letter of the law relative to PA 102 (2011) was scrutinized. By identifying the outliers, the authors contended that truthfulness can permeate the study. Having staff consider their own data and reporting on their own conclusions provided sound analyses and interpretation.

Consideration as to the participants involved was designed to reflect employees currently serving in the district as well as former employees. The researcher is currently the superintendent of the district where the research was conducted. The research is limited by the omission of state assessment data as information relative to what the state is utilizing is not readily available, nor has the baseline for establishing proficiency of state assessments been shared with the local districts.

The lessons learned by the district could provide great utility to the State of Michigan in addressing future legislative mandates and initiatives. The critical theory of research cited by Carspecken and Walford (2001) contended that findings are useful if the results liberate the individual or group studied, and if the information learned provides meaningful changes in how they deal with the phenomenon. These criteria are facilitated throughout the study and at its conclusion. Yin (2003) identified six sources of data that can be infused into a case study. They are (a) documentation that can be readily reviewed, exact details incorporated and not created as a result of the case study; (b) archival records that are quantitative and precise; (c) targeted and insightful interviews that focus on essential areas of the study; (d) direct observation in real time; (e) participant observation with identification of interpersonal
behavioral motives and (6) insight into physical artifacts. This study aimed to incorporate all of the sources of evidence suggested by Yin (2003), save the observation component, due to the potential manipulation of the researcher’s involvement in the study.

**Implementation of Study**

The protocol of the researcher was the following:

1. Seek approval of Eastern Michigan University’s Human Subjects Review Committee (see Appendix A).
2. Request participation of ten teachers (5 retired and 5 members of the professional negotiation team), three principals (no longer with the district), the former superintendent, the former board president, the current board president, and technology director (see Appendix B). Collect and review data of the (10) teachers to be studied from the Stages Software database.
3. Collect and review district artifacts relative to InFORMED and the evaluation tool as a whole.
4. Conduct interviews with identified stakeholders (see Appendices C, D, E, F).
5. Ask participants to validate data through member-checks.
6. Analyze data and report findings including identifying the adopter types as highlighted by Rogers and Shoemaker (1971) in an effort to answer what the district did to prepare for the integration of PA 102 (2011), what strategies were integrated for sustainment of the evaluation tool, and unveil what the district has learned from the change initiative.
Validity and Reliability

The researcher was aware that he has had a considerable role in the process the district has experienced relative to the evaluation of teachers, as he has served as a principal and as superintendent during the period studied. The research is limited by the omission of state assessment data as information relative to what the state is utilizing is not readily available, nor has the baseline for establishing proficiency of state assessments been shared with the local districts.

Legal, Ethical, and Moral Issues

Organizational and individual influence was considered by the researcher in this study. Choo (2006) identified that organizations influence in their actions and talk which, according to Brown (1989), translates into active cultivation of establishing thoughts relative to schema and understanding the view of the world where the phenomenon exists, Sergiovanni (2006) said that districts share some common characteristics that are uniform; however, Choo (2006) believed that organizations seek to exert their own scope of influence. Earlier, Daft, and Weick (1984) opined that schools could be categorized as loosely coupled organizations. As such, identifying opportunities for individuals to share their interpretation on the district’s approach to conforming to PA 102 (2011) was beneficial to protect against the portrayal of the researcher’s opinion being the exclusive representation of the reaction of the district to PA 102 (2011).

The Human Subjects Review Committee of Eastern Michigan University reviewed and approved the proposal for this study prior to beginning any research or contacting any participants (see Appendix A). Permission to utilize the data and the district as a case study was granted by the board of education president. In relation to ethics, Bogdan and Biklenen
(2003) conveyed that it is necessary to have subjects enter into the study by their own accord and of their own volition. During the onset of the invitation to participate, subjects were advised of the objectives inherent in the study. Bodgan and Biklen (2003) also suggested that participants be afforded a means to withdraw from the study at any time. Effort was exerted to provide sanctuary for participants from any potential risk. In describing the district and personnel, pseudonyms were distributed, and after data were analyzed and coded, they were destroyed after a reasonable period of time, as suggested by Creswell (2009).

To ensure validity and reliability, authenticity was a considerable objective of the study. The researcher was cognizant of the fact that he had a significant role in the process the district has experienced relative to the evaluation of teachers, as he has served as a principal and as superintendent during the period studied. The potential influence of the researcher’s involvement in a study of this nature is noted due to the inherent role he has played in the organization. Schwandt et al. (2007) believed that a considerable component of ensuring authenticity is to integrate fairness throughout the study. It is suggested that authenticity is achieved by integrating overt and collaborative descriptions of the findings regardless of what the researcher finds. Schwandt et.al. (2007) further advised that member checks are implemented where participants have a forum to review the findings to validate authenticity.

The researcher was cognizant that, as superintendent, the potential to exert bias existed. A statement to participants of the desire to improve educational outcomes through improved evaluation systems was distributed to participants of the study.
Data Analysis

The researcher utilized reflective analysis, as a considerable amount of the data was able to divulge both surface features of the phenomenon of teacher evaluation and the true essence of what PA 102 (2011) procured both intentionally and unintentionally. The consequences of enacting this law told a story of the implications of such dramatic alteration in the evaluative landscape. Based on interpretivist epistemology, as defined by Gall et al. (2005), the researcher believed that each participant and stakeholder would have a unique impression of what role that the teacher evaluation tool played in the educational operations of the district studied.

To ensure the credibility of the study, the researcher integrated outliers as recommended by Miles and Huberman (1994) and compared and contrasted the findings to the main-stream responses procured by the researcher. Self-reflection was utilized to identify patterns among the data. The potential impact of the researcher being the current superintendent of the district was mitigated by the inclusion of two research supervisors with teacher evaluation clinical experience and expertise as central office administrators and building principals.

Participants of the study were identified following a review of their responses in accordance with the prospect that each participant would be characterized as one of the following, as described by Rogers and Shoemaker (1971): (a) innovator: eager to try new ideas, open to change, willing to take risks, and usually perceived as naïve or a little crazy and therefore not well integrated into the social structures; (b) leader: open to change, but more thoughtful about getting involved, trusted by other staff, and sought for advice and opinions; (c) early majority: cautious and deliberate about deciding to adopt an innovation;
tends to be a follower, not a leader; (d) late majority: skeptical of adopting new ideas, and “set in their ways;” can be won over by a combination of peer pressure and administrative expectations; (e) resister/laggard: suspicious and generally opposed to new ideas, usually low in influence and often isolated from the mainstream.

The basis upon which this framework is devised hinges on the notion that there are very few innovators and a small number of leaders and resisters, but the majority of participants in an organization could be described as early majority or late majority. In reporting the findings, the role each of the participants was noted. In addition, the study yielded information about the impact that the district’s process for implementing change relative to PA 102 (2011) had on determining if administrative support and approval was sufficient to support the sustained change needed to benefit students, as suggested by Hall and Hord (2015).

**Summary**

This chapter served to identify the research questions and the tradition in which the study was conducted, with a particular emphasis on case study inquiries. The setting for the district was reviewed so as to provide a contextual reference point. Research methods relative to interviewing stakeholders were incorporated in conjunction within the implementation of the study. Validity, reliability, legal, ethical, and moral issues were also addressed in this chapter. A protocol for procedures, collection of data, and the analysis of the data were also discussed in this chapter. The importance of the findings for the district and the potential impact that the study could have for the district, school leaders, and the field of education, including policy, were also addressed.
Chapter Four: Findings

Introduction

Dramatic changes in teacher evaluation have transpired since 2010. The purpose of this study was to identify how one district attempted to meet the need for greater accountability in the teacher evaluation process as mandated by law and to describe strategies that were employed by the district to ensure the integration of student achievement data.

The chapter begins with a brief review of recent changes in teacher evaluation policy in Michigan, followed by a discussion of Rogers and Shoemaker’s (1971) adaptor model that identified five categories into which individuals within an organization fit when innovation is introduced. Data gathered in open-ended interviews are analyzed, and emergent themes are identified.

Evaluation Policy Overview

Incorporating student achievement results into teacher evaluation was advanced in 2011 when the Michigan legislature reformed the Teachers’ Tenure Act in response to federal policy (Act No. 100, 2011; Act No. 205, 2010; Heitin, 2011; McGuinn, 2012). State teacher evaluation legislation now requires teachers to be evaluated annually based on multiple classroom observations (Act No. 102, 2011; Act No. 205, 2010). In addition, for each instructor, teacher evaluations are to result in the summation labeling of ineffective, minimally effective, effective, or highly effective (Act No. 102, 2011). Public Act 102 (2011) included the requirement that a temporary commission referred to as the Governor’s Council on Educator Effectiveness (later called the Michigan Council for Educator Effectiveness) be established to review and recommend the most appropriate tool. Legislation (Act No. 102,
2011) charged this council with identifying a teacher evaluation instrument to be exclusively implemented in the State of Michigan.

The tool was to incorporate “instructional leadership abilities, teacher and pupil attendance, professional contributions, training, progress report achievement, school improvement plan progress, peer input, and pupil and parent feedback” (Act No. 102, 2011, n. d.). Although numerous options were recommended by the Michigan Council of Educator Effectiveness (originally called the Governor’s Council on Educator Effectiveness); Executive Summary, 2013, legislators have not officially designated a state framework or data tool to use. The options included: Charlotte Danielson’s framework for teaching (2007), the Marzano teacher evaluation model (2007), the Thoughtful Classroom (Silver Strong Associates, 2007), and 5 Dimensions of Teaching and Learning (Center for Educational Leadership, 2016).

The district being studied had already implemented the Danielson framework (2007) at the time the statute was enacted; however, the ranking of educators and inclusion of student achievement data had not been done. The district being studied was one of the first districts to aggressively integrate the mandates inherent in PA102 (2011).

**Rogers and Shoemaker’s Adaptor Model**

Rogers and Shoemaker (1971) identified five categories into which individuals within an organization fit when innovation is introduced—innovators, early adopters, early majority, late majority and laggards. Not everyone has the same level of responsiveness to change and the externalities associated with change. The ability to be responsive is the main dependent variable in the sequence of research relative to change. Innovators are considered venturesome and eager to try new ideas, typically have access to large capital resources, and
embrace the concept of risking potentially losing against the potential reward of winning big. Innovators do not necessarily garner the respect of those they lead, as they can be viewed as brash; they must possess the capacity to deal with setbacks. Most certainly, Joan as the former superintendent fit this component almost completely. Rogers and Shoemaker (1971) also noted that the ability of an innovator to launch a new idea into a social system by importing innovation from outside of the system’s boundaries also provided the climate for the change initiative to ensue. The business background with which the superintendent and board members came to the table provided the initiative to embrace the change; this coupled with the authority of the law provided a setting where rapid implementation was possible.

Early adopters are considered to be widely respected. All three of the principals interviewed appeared to fit into this category. John, Anne, and Rod all spoke repetitively of the importance of establishing trust of the staff. The principals were more integrated into the social system of teachers than may have existed in other districts, as they were working with staff on a process of evaluation that consisted of pre- and post-conferencing. A pedagogical review was the focus of the Danielson (2007) instrument.

Potential adopters seek the opinion of the early adopters and the respect with which these individuals are vested leads others to invest their trust in the opinion-centered leadership of the early adopters. John, Anne, and Rod seemed to serve as role models for the early majority. The impact of union member, Bob, who also appeared to be an early adopter also could have played an important role in the advancement of the evaluation process and its early adoption.

People in the early majority category typically adopt the new idea just before the average adopter. It appeared that the teacher leaders who were involved in this study
represented this population; that group does not lead, but follows with intentional willingness. Kristina, as technology director, also appeared to be assigned the designation of early majority. She understood that the changes were coming, and that her job dictated that she must attempt to construct a technological medium to track student achievement. Three teachers—Krista, Brooke, and Fred—all appeared to share common characteristics with early majority. They concurred that the evaluation process was part of determining a more conclusive reflection of what constitutes effective instruction.

Late majority adopters adjust slightly after the average member of a social system. Of the ten teachers identified in the study, it appeared that six participants could fall within this range based on their interview responses. Adoption becomes a necessity due to networking pressures. According to Rogers and Shoemaker (1971), the late majority group approaches innovation from a skeptical and reserved perspective and don’t fully invest until most others have done so. The influence of the social norms must be fully exerted before acceptance transpires. The late majority would prefer that all uncertainty is dissipated prior to adoption. It did not appear that regarding teacher evaluation this would ever be completely attainable to achieve.

Laggards are considered traditionalists. They have almost no opinion about leadership in the change initiative and would prefer to have the framework for decisions based on the precedent set by previous generations. None of the participants in this study were identified as laggards.

**Participants in This Study**

In this qualitative case study design, 17 participants were selected based on purposeful sampling that included the participant’s current role in the district and past
responsibilities. Interview responses of the participants in this study revealed how each made sense of the evaluation process in the context of the school system. Their views reflected what the researcher believed to be each stakeholders’ relative social system adaptor position according to Rogers and Shoemaker (1971). Table 1 shows the participants by role and gender.

Table 1

*Study Participants by Role and Gender*

<table>
<thead>
<tr>
<th>Role</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Administrator</td>
<td>2</td>
</tr>
<tr>
<td>Technology Director</td>
<td></td>
</tr>
<tr>
<td>Teacher Leader</td>
<td>2</td>
</tr>
<tr>
<td>Former Teachers</td>
<td></td>
</tr>
<tr>
<td>Board of Education</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2 introduces participants by pseudonym and adopter category, as determined by the researcher.
Table 2

Participants by Role, Pseudonym, and Perceived Adopter Category.

<table>
<thead>
<tr>
<th>Title/Role</th>
<th>Pseudonym</th>
<th>Adopter Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Former Superintendent</td>
<td>Joan</td>
<td>Innovator</td>
</tr>
<tr>
<td>Former Principal</td>
<td>Rod</td>
<td>Early Adopter</td>
</tr>
<tr>
<td>Former Principal</td>
<td>John</td>
<td>Early Adopter</td>
</tr>
<tr>
<td>Former Principal</td>
<td>Anne</td>
<td>Early Adopter</td>
</tr>
<tr>
<td>Former Board Member</td>
<td>Clark</td>
<td>Innovator</td>
</tr>
<tr>
<td>Current Board Member</td>
<td>Buck</td>
<td>Innovator</td>
</tr>
<tr>
<td>Technology Director</td>
<td>Kristina</td>
<td>Early Majority</td>
</tr>
<tr>
<td>Former Teacher</td>
<td>Mike</td>
<td>Late Majority</td>
</tr>
<tr>
<td>Former Teacher</td>
<td>Krista</td>
<td>Early Majority</td>
</tr>
<tr>
<td>Former Teacher</td>
<td>Brooke</td>
<td>Early Majority</td>
</tr>
<tr>
<td>Former Teacher</td>
<td>Fran</td>
<td>Late Majority</td>
</tr>
<tr>
<td>Former Teacher</td>
<td>Julia</td>
<td>Late Majority</td>
</tr>
<tr>
<td>Teacher Leader</td>
<td>Melania</td>
<td>Late Majority</td>
</tr>
<tr>
<td>Teacher Leader</td>
<td>Fred</td>
<td>Early Majority</td>
</tr>
<tr>
<td>Teacher Leader</td>
<td>Bob</td>
<td>Early Adopter</td>
</tr>
<tr>
<td>Teacher Leader</td>
<td>Tim</td>
<td>Late Majority</td>
</tr>
<tr>
<td>Teacher Leader</td>
<td>Mikayla</td>
<td>Late Majority</td>
</tr>
</tbody>
</table>

Making Sense of What the District Learned from Evaluation Changes

Former superintendent Joan. Superintendent at the district being studied for 10 years from 2001 to 2011, Joan earned her doctorate degree from Central Michigan University (CMU) in educational leadership in 2011 and subsequently left this district to assume the role of superintendent at an intermediate school district. Earlier, Joan earned a Bachelor of Science degree from Ferris State University and served as an assistant superintendent of business at a larger suburban district after completing a master’s degree in educational
administration from CMU in 2000. Since 2015, Joan has served as a union director for the Michigan Education Association (MEA). Joan has never been a certified teacher.

I met with Joan at her home to discuss her recollection of the evaluation creation, integration, and protocol that was initiated under her leadership. Prior to her arrival, the district didn’t have a strong system of accountability for teacher evaluation; generally, the principal conducted the evaluation. The interview transcript provided most of the data; artifact review information was added as appropriate and noted accordingly:

When I first came to the district, teacher evaluations were supposed to be done every three years. It was a yes, no, or satisfactory/unsatisfactory protocol. It was two or three pages long from my memory. Very subjective to the evaluator, and there was no standard practice, procedure for how it was done other than the timing. It was satisfactory or unsatisfactory and… some teachers hadn't been evaluated in the 25 years that they taught there…some teachers were evaluated, but it depended on the principal.

This summation offered by Joan seemed to permeate the research and provided a theme that the district had not engaged in a specific, systematic evaluation process, such as indicated in the transition to a model of integrated evaluation process with PA 102 (2011). A review of all the study participants’ personnel files indicated that the teachers had been given narrative evaluations called *Teacher Appraisal Record* from at least 1977 until 1999 when a document titled *Teacher Evaluation* used from 1999–2004 was introduced that allowed three potential ratings—satisfactory, unsatisfactory and not observed. The instrument had three domains: (a) subject, which encompassed knowledge of teaching, methodology, and evaluation; (b) management, which comprised organization and care of room and equipment;
and (c) relationships, which differentiated between self-relationships and interpersonal relationships. In 2005, the district made the transition to the Danielson framework (1996) for teacher evaluation.

Joan expressed that outside of the timing and the negotiated evaluation instrument that was bargained by the teachers’ union, the district had not pursued teacher improvement through evaluation systems prior to her administration. Joan perceived that prior to 2001, a sense of complacency existed, and mediocrity was acceptable. Joan pointed to the lack of an individualized development plan (IDP) program for new teachers as an exhibit to her point:

Individual development plans were not done for new teachers. It was, if they were struggling, the principal would work with them, but pretty much everyone was considered effective or adequate. I'll use the word adequate. Everyone was considered an “okay teacher.” I don't think that they dismissed any teacher from my memory. Before my time I don't think anyone was held to any standards other than their own intrinsic or internal measurement of, you know, good teaching.

A review of teacher personnel files indicated that Joan had been partially accurate in that new teachers did not start receiving IDPs until fall, 2005. It was in the spring of 2005 that Joan worked with one of her administrators on overhauling the entire evaluation process at the district for the first time in her tenure:

I was overseeing a master program. Will Kearney was the assistant principal in the high school, and he was doing a master's program and part of his program required him to assess the evaluation protocols. Through that process I learned about the Danielson (1996) rubric, and at that time we negotiated with the teachers to implement the Danielson rubric up throughout the district. So that was like a two-
The first year we did it we... We got training, so Mr. Kearney and I obtained training, went to training about the Danielson rubric, and then we trained the other administrators. We brought in Larry Audette for a 5-day training, and he did the whole shebang.

Joan felt that there was a lot of training done at the time the district shifted towards the Danielson framework in 2005. She seemed to convey that this helped later in the transition to adjusting to the ramifications of PA102 (2011). The creation of common language seemed to be a piece that Joan felt strongly about, as she wanted to return to the Danielson piece of the instrument:

The good thing for us is our employees had, I think, about four or five years of using the Danielson rubric to understand it, to know how it works. We as leaders... I remember being in a meeting with our leadership team and talking about what really is a high performing, a highly effective person. What does it look like? We built a language rubric around just what that is and the indicators for that in our district...in 2010, when the law did change, it was the law. And the board at the time required us to follow the law—all aspects of it. If it's the law, make it happen. So the administrative team and I worked together.

Several brainstorming days, and meeting, and thinking about how it could work, how could this happen? And rolled out a plan that we thought would work, negotiated it with our union, and implemented when it was mandated. They ended up then extending the deadline, but we had already implemented. There was... I wouldn't say that it was completely smooth. There was a lot of angst, but... I would say 20% got on board, and teachers helped us.
Trust as the contributing factor, according to Joan, explained what she seemed to convey was a large percentage of support amongst teachers that was garnered for the rapid implementation and integration of student achievement data into the teacher evaluation tool. In regard to moving forward, teacher support, in conjunction with the demand from the board of education to comply with the law as promptly as possible, appeared to give the district latitude. Joan believed that having relationships in place in conjunction with having multiple measures for student growth besides solely the MEAP as the data piece was critical to instituting student data that would contribute to the overall picture of a student’s growth and subsequently a teacher’s effectiveness:

We believed that the MEAP test was not an accurate assessment of the students' knowledge or learning. It was one piece of data, one piece of the picture, but each child has a learning puzzle, in my mind. And that learning puzzle has lots of pieces that are moving. They're not stagnant. They ebb and they flow. The MEAP was one piece; their grades were another piece. Their attendance was a piece. There's so many data points for a child that impacts learning. And we knew that we did not have...

Although we had common assessments and pacing guides, we did not have objective assessment data, and so the district, the board committed to buying Northwest's evaluation assessment, which was an objective, standardized, adaptive assessment of where a child is, and that became, I think, a critical data point that kinda tied all the other pieces together.

Having the various data points available was a piece that Joan felt was essential. However, she would regularly return to speaking about the value and necessity associated with the change initiative to be inclusive of trust and relationships. Joan expressed that there
was a concerted effort to tie herself and all administrators to the student data right from the start to promote a sense of unity among the administrators, teachers, and students:

They could have been resistant at times, and it was my job to create the impetus, or to not let that mindset stop us from accomplishing and getting to the goal, and doing it with extreme fidelity and extreme... intentionality to do it the best we could, but again, the team was great. I don't know. This probably couldn't have happened in a different district with different people. This is not about a me thing. This is about the team thing. It happened because of every person on that team… I thought it would do something, because it was looking at efficacy of instructional practices, instructional curriculum, instructional delivery in a different way. It made educators, the system, the organization take ownership in a way that did not exist before.

Joan felt that the district was successful at building the tool and integrating it. However, in reference to Joan’s opinion in regard to the effectiveness of how the district created the evaluation tool, she felt compelled to convey that she had been numerous other places, and her ability to reflect is tempered by what she experienced after leaving her role as superintendent of the district being studied in 2011:

I think the performance in the year and two years following the implementation of linking data to evaluations shows that it made a difference, and students did better. Now, seven years later, I'm not in touch with the district, so I don't know. I hear that they're doing very well. So do I think it was a good thing to do? Would I do it again? Yes. It was worth the heartache, it was worth the battle, it was worth the disruption in the norm. It was worth it.
On the flip side, you have to have trust, you have to have... You have to build in an evaluation system first before you try to do this, because it's too scary and there are too many... So, when I left the district, I've had the opportunity to work on the other side of the table and represent some teachers and with administrators who are heartless, careless power-mongers that are not motivated by helping students learn. They're motivated by wielding power and squashing human beings, and that's distressing. It's disgusting, and I've seen the other side.

Joan seemed to indicate that after experiencing other environments and perspectives, she has come to reflect that trust is the key ingredient to successful navigation of an intricate and complex system within an organization. The scope of the district being reviewed was the formation of relationships predicated on trust. The assessment that Joan offered is a theme discovered throughout the inquiries conducted with other stakeholders within the district.

It is worth noting that Joan indicated a tremendous amount of battle and she even carefully places emphasis on the term heartache. Her reflection suggested that not everyone was an early adapter to the initiative that was mandated by PA102 (2011) or the processes the district created and implemented. Joan offered:

The data piece was tied to the students, and it’s, “Hey, we're in this together” type... it only makes sense to blame the parents for so long, it only makes sense to blame genetics for so long, it only makes sense to blame the kids for so long, but ultimately it made educators, and I don't mean just teachers, I mean educators: teachers, principals, board members, community. It made us all responsible, or to a degree, responsible for the learning. I'd like to point out, though, getting us there was a, at some times, a difficult mind shift for everyone to think about.
And there were sometimes when I had conversations with leadership...I remember specifically, a conversation that was, “This teacher is performing poorly.” The principal knew they were performing poorly, but, “They're not that bad.” Well, yes they are…if the leadership stays committed to it, if the process continues to be honest, trusting, transparent, kind, caring. Take your ethics. You have to do it with all of the ethics care or critique, justice. You have to. Because if any one of those pieces is missing, it won't be believable. If it isn't real accountability and you don't really hold people to the standard and you give them a pass, then it becomes meaningless. Right? That's the justice. Care and critique, you have to do it with the right heart and the right mindset.

Joan did not believe that the state effectively navigated the change initiative but that the district being studied jumped on the data piece and saw gains only after the student achievement was connected to the evaluation instrument. The ensuing collective data growth of the district did not transpire when the Danielson (1996) part was initially invoked, but the introduction of data used in tandem with that modified Danielson qualitative component bore light onto academic growth in the district:

In the state of Michigan, extreme failure because the state didn't take leadership, and they still aren't. They still won't. There's warring factions. They don't have anybody to stand up and say, “I'm a champion, and here's how it is. And look at this district did it, and this district did it.” I mean, yes, they're promoting that, but they're not making anything state-mandated. And so, how many districts, more than half haven't even implemented anything, and they're not gonna until they are told what to do and how to do it. Okay. So that's it.
At the state level, it's a failure. It's still a failure. So, I think they're 50-50 [Danielson and student data]. I think that they're mutually beneficial. I don't think that you could really implement. You could implement an evaluation tool without data, but when you do that you may not see the results because there is no... It's like the rest of the story, right? There's no picture. You get halfway through the story and you're just at the just happened, and then what? It's like stopping midway. And we'd done that for 20, 30 years. It obviously didn't work.

Former principal Rod. Rod served in the district as assistant principal or principal for four years when the change initiative associated with PA102 (2011) was enacted. He served two more years as a principal working within the framework that the district integrated. Following the departure of the district, he served as a principal in a public high school just outside of the state capitol. Rod earned his bachelor’s degree in secondary education from Western Michigan University in 1989 and his master’s degree in educational leadership from the University of Cincinnati in 2007. Rod served for more than 20 years as a teacher prior to becoming an administrator in a rural Michigan district, and acted as union president for numerous years while a social studies teacher. Rod no longer has any association with the district in this study. Rod retired in 2016 but is contracted back by another district on an annual basis to complete 40 teacher evaluations on behalf of the district using the 5 Dimensions of Teaching and Learning mModel for teacher evaluation.

At the inception of Rod’s introduction to the district in this study, the modified Danielson rubric was being utilized and all teachers were not evaluated on an annual basis in 2008. Rod spent time describing the processes and the evolving nature in regard to the
evolution of the evaluation protocol to which he was introduced. Rod was interviewed at a hotel conference center for the purpose of this study:

So that first year, I don't know... Maybe 15 evaluations I did. That was true in the next school year as well. '09-'10. We would stay for the entire hour. And they were required to be observed twice. I can't remember all the requirements, but yeah, we were in the classrooms twice. So, for those teachers, and then you write up one evaluation at the end. Evaluations were announced. So they knew when we were coming. It was set up with the teachers, what hour they wanted us to observe and what not. We introduced the... I can't even remember what it stands for, NWEA...

North West Education Assessment. I think is what it was, as our growth model, which was administered in the fall, in the winter, and again in the spring. The goal being that you had to show..., students had to show a year's worth of growth with a particular teacher. For instance, a math student; For the math teacher to be effective, it'd show a whole year's worth of growth. We used the reading for those areas, social studies, music, that kind of thing where the teachers didn't have a specific task... social studies didn't have a test item. So they used a reading score, and had to show a year's worth of growth in reading. We also came up with the InFORMED data rubric in conjunction with that...and that was a system where we quantify...Danielson [1996]. One, two, three, four points for each of those, four points being best. So if you were a distinguished teacher, you earn four points. In the fifth domain was the data. There was a numeric point system identified.

Rod recalled the process of creating the system as being exciting due to the nature of the team who designed the tool being implemented in the district rapidly, shortly after the
inception of the law. He recollected with laughter at the initiative that the team took in
tackling the objective of making sense of the changes in legislation and how it could be
quickly integrated into the systems of the district:

The design was the administrative team, superintendent, all the way down. We were
not very common, or we were not very... What's the word? People didn't like us at
conferences 'cause we would walk out of what they were talking about, and go do our
own conference in the lobby of the Bavarian Inn...we did some of the work right
there to create it... we take their knowledge and go out and put it into effect and
created our own system. That's how we came up with it. It was nice.

Rod was even more emphatic, as he credited a large degree of the improvement that
the district incurred to the contribution of the improved evaluation system. He claimed that
the assimilation of the data forced a more concerted mindset to the Danielson rubric (1996),
which was a demonstration of quality instruction. He claimed that the teachers had a greater
desire to be concerned, as they were now directly connected to quantifiable data and this
created a greater sense of urgency to consider how to move within the framework of
Danielson and InFORMED:

... and greater accountability. I think that there was a greater awareness if nothing
else. Just a greater awareness of the elements of the Danielson rubric [1996], and a
concentration on how do I move from basic to proficient, or proficient to
distinguished. Lots and lots of those conversations of, “I don't like being basic. What
do I have to do to be proficient?” Or “I don't like being proficient, I think I'm
distinguished.” So conversations centered around how do you move up the scale so to
speak. The staff trusted me to tell them what they had to do. Many were cognizant,
and asking, and looking for input as to how they could get better, and “is this what you mean?” would be a common conversation.

I'm thinking of my last year. I know there was a definite concentration on the purpose, the learning objective, whatever word you want to put in there, where the buzzword now is learning targets, but that really focused on, “This is what we're gonna do. This is why it's important, and this is what it looks like to be successful.”

So I think, yeah, absolutely. The tool and process had to result in better planning, better organization, and better teaching as a result of it. A switch from teaching to the old idea of, ‘Well, I taught it” to “Did they learn it?” I think is the big change. Despite the overall perceived belief that the process harvested growth and improvement for the district, Rod was quick to acknowledge that there were skeptics within the teaching ranks, but it did not deter the administrative team from forging ahead:

There were always skeptics, but I would say the vast majority of the teachers were...willing to accept this change. The administrative team at the district, I think we had... I think as a team, we taught each other very well. I think we had a grasp. It was a little more challenging to get staff onboard for this groundbreaking work. We were asked to speak across the state on what we were doing because we were a forerunner. We were implementing the law while others were looking for a loophole to the law or just simply deciding that they didn't have to follow the law.

Upon learning that the district had implemented a system, Michigan Education Association (MEA) contacted the district and wanted to confer about the instrument and process being enacted. Rod recalled that the administrative team met with Frank Ciloski as the MEA vice president. Rod recalled the meeting with apparent pride:
We walked in and of course they were opposed to what we were going to do. By the time the meeting was over with, we were side-by-side with them. They were onboard when we explained what we were doing, why we were doing, how we were doing, what the teachers had... The opportunities that we were giving teachers. We were endorsed by the MEA, and really, it took off after that.

Rod pointed out with affection the numerous venues where the administrative team were asked to provide leadership and help other districts integrate a system of their own. He expressed that the district being studied was so far out ahead and completed such a large amount of work so rapidly that administrators were considered experts amongst peers across the state:

We were very much a leader across the state. I'm sure that Zimco has benefited financially from us pushing it and us sharing with other districts and big districts. We talked to Saint Clair Intermediate Schools, we talked to Ann Arbor Public Schools, we talked to the... the Michigan Teacher Evaluation Conference held at the Lansing Center.

The integration of the evaluation process became easier after meeting with state union leadership. Rod relayed that the process within the district became easier, and suddenly administrators in the district were promoting what the district had done. Rod conveyed that the endorsement of the MEA resulted in immediate notoriety:

It definitely knocked down the barriers. When the union is not only not going to fight you, but is endorsing your plan, yeah, you become popular…I knew it would have an impact. You got to remember, 2008, 2009, 2010, those were assaults on education with the loss of teacher tenure, insurance was changing, seniority was changing.
There was just...we were being assaulted right and left by Governor Snyder. So when this teacher evaluation thing came down, yeah, it probably was viewed as an assault on teachers rather than the tool that it's become for growth.

Trust can't be underplayed in this at all, especially given that political climate that I just talked about, because teachers don't trust anybody, rightfully so…the whole profession was really just being hammered, so, absolutely there's a fear, and absolutely the teachers played a huge role in. Again, there were skeptics, but for the most part, there was a willingness by most of them to give it a try, let's see how it works. We've got to do it, so bless their [teachers] heart, it was, “Okay, if we're gonna do it, let's do it the right way.” I think that our own union members were on board with it, worked with their membership to allay any fears that came up…it would be nice if teachers really saw it as, not a tool of judgment, but really as a tool of growth.

**Former principal, John.** John began working with the district in 1977 as an elementary teacher and was later promoted to principal, where he served for 16 years. In that time he was a principal at all levels of the district being studied (elementary school, middle school and high school). At the time he retired in 2012, he was serving as high school principal. John took one year off in retirement and has since accepted a position as director of categorical funding. In this capacity, part of his responsibilities include evaluating all federally funded teachers and programs. He achieved both his bachelor’s degree in elementary education in 1977 and his master’s degree as a reading specialist from Central Michigan University.

John was interviewed at his home. John provided a deep historical perspective about teacher evaluation in the district in the years preceding the enactment of PA 102 (2011). He
reached back to recollect what he experienced as a teacher at his hiring in 1977 and walked through the procedural evolution. John was direct when pointing to improvement process:

When I was classroom teacher, the negotiated evaluation tool was pretty much subjective based on observation…it was part of the teacher contract. I don't remember, specifically, if it was one or two, but that observation narrative was provided, and so, that tool as I said, was pretty subjective…even though it may have had an overview of a whole year, there wasn't a lot of required documentation required. There was a formal observation notification type of process involved, but as far as a rubric, as there is currently, there was nothing really, that addressed specific pieces of an evaluation tool… I would say there have been several areas of improvement, even though over time, the evaluation tool may still have been part of the teacher contract, and discussed and addressed in that realm. I think the improvement part is previously, there really wasn't... As a teacher, you may have identified a goal for improvement, however, there really wasn't any teacher or teaching rubric that was identified. So it was kind of up in air, if I was the evaluator, then it was whatever I saw, observed, or felt should be, based on. Maybe just my perspective, as opposed to having a set tool with detailed rubric as a current means. The other part of that, I guess, is, having a more finite rubric to look at, I think allows more specific discussion, focus; looking at strengths and weaknesses also allows discussion and opportunity to build relationship between administrator and teacher based around that evaluation tool.

John was pointed in directing that the establishment of the moments when conversations that were allowed to transpire as a result of working through the Danielson
rubric (1996) and the InFORMED component transpired is when growth ensued. He highlighted the relationship(s) being cultivated among teachers and administrators as critical. John’s perspective of having worked through the entire progression of 40 plus years in the district as an educator and a principal led him to the conclusion that prior to 2010, the emphasis on student achievement had not been ingrained with the emphasis that student achievement would eventually demand of individual teachers. John noted that the data as a collective whole was scanned at a macro-level:

Going backward from the most recent history would have been MEAP test, so that would have been from the early ‘90s, up until M-STEP took over. Yes, that was looked at in the more recent times, and as teacher evaluation did move into more structure that was a tool that was able to be used and analyzed. However, it wasn't always directly impacted; it wasn't as a teacher, or your classroom data, wasn't really there. Back in history, California Achievement Test was used. That data was looked at, but I don't believe it made connections to...how you were evaluated, or that tied in to your teaching, the strategies you used, or direct link into evaluation… I think the process was already in place, but in the current times, most definitely, that data is being looked at, tied back to school, grade level groups, teacher absolutely, and trying to determine whether they are the best teaching strategies that can be used to help students become more successful.

John believed that the Danielson integration at its inception began to formulate individualized dialogue and discourse aimed at improved instruction, but it was the quantifying of the tool and the student achievement that made the difference in the process
becoming more persuasive. John also seemed to allude to the conditions associated with teachers striving for improvement as a result of the ranking associated with the legislation:

When we leave and talk about the current tool that we have, Danielson (1996), I don't think it was specific globally. I think it was in a one-on-one setting. You could identify strengths that a specific teacher, when you're talking to the teacher, and as an administrator, I obviously had a better global perspective of the group of teachers that I worked with, and who may have the most strengths, and which people have weaknesses. But I don't believe, initially, that information was shared, or used as a determining factor to, rank teachers or identify who are the best teachers in the district.

John felt that the relationships that were built through quantifying what quality education is among the principal and teacher while comparing data and the Danielson rubric (1996) can’t be lost in his account. Part of the catalyst to this growth is described by John as working through challenging the data with which some teachers would be credited, as he had initially had to hand calculate the immense volume(s) of data:

I remember being at the high school, when I moved to that realm, and we were looking more at data, and time data, specifically to teachers, and it wasn't automated—in trying to do some of that manually, and making mistakes with teachers, and telling them that, “This might be an area that you're not as strong as in other areas,” and then, having recalculating student data, and changing, turning that around, that challenge, I guess, also allowed... Even though it was a challenge, the positive part was that it was part of that building relationships with teachers, in the fact, I'm spending time talking with them, in the sense of pre-observation, post-
observation, the time you spend together talking about that. Even though it was time consuming, it was a very positive part of that process.

I still think that an important part of it is the time spending with teachers, because I think you can pick up snippets of what's going on in class, being in and out short-term, as opposed to a lengthy formal observation, I think you can see just as much being in and out, whether it's a walk-through formally, or just walking through the building, but I think that an important part still is that time you spend with teachers talking, sharing...the relationship most definitely. I think like anything, it's evolved, because I think until teachers and other administrators had a better understanding of that process and utilizing the components of the evaluation, following through on that, talking, sharing and even making specific recommendations to teachers based on the rubric. Until you do that, it doesn't really mean anything, so I think that's part of that processing.

John delved into the process of rolling out the evaluation and referenced how staff members were critical of the promotion of the tool being integrated. John referenced how part of what the district recognized was that positive components of the instrument emerged as the process and integration became more congruent:

The tool contributed to a higher degree for student achievement, as well as school improvement, because I believe that people have had the time to learn and understand the details of the rubric, what they mean, how they're interpreted, the more we've talked about it; then people come closer to consensus. Everybody, like anything, may not agree 100%...we've been using this tool long enough that I think people are closer on their understanding of what the tool means, and as a result of that, going back to
teaching, learning, strategies—what our best practice is, focusing on specific components.

The rubric has the evaluation and has allowed people to work together to get to that...I think there's always some hesitation when there's change and something new, but at the same time, it's somewhat of a plan. The evaluation truly is based on research for what is quality teaching, what are the characteristics of quality teachers and the actions they're teaching that affect student learning. So, once you get past that initial, “Oh, we're just being told to do this,” but nobody says the details, because the evaluation's based on research, then you also have that as your backup to know that these are the quality components, and so, this is what we're looking for. Those quality components of teaching also affect student learning, so I think that's what the process is based on. In the realm of school improvement, we've always had, what are “research-based strategies” that you want to have in the big picture; well, it kind of ties directly into that inspires that process.

John emphasized the importance of the district integrating an evaluation process that was embedded in research, considering that the Danielson (1996) component had been integrated into the district for numerous years prior to the introduction of the student achievement piece, the Fifth Domain or InFORMED. Acknowledgment of incorporating the Danielson rubric that had already been researched and tested as a central piece of the evaluation process led to conversations between John and his teachers about quality instruction. John considered the Danielson rubric and the dialogue with educators individually and collectively that ensued in the review of the student achievement portion of the instrument led to key factors in the district’s growth and academic acknowledgment that
the district experienced. John indicated that prior to the conversations and the emphasis on ranking, data and identifying which data to look at weren’t necessarily a central focal point on which teachers concentrated. John expressed his belief that the tool narrows the focus and that the rubric identifies quality instructional practices:

More and more people are beginning to make that connection that the data is a much more important part than it ever has been and that the data and the evaluation, the narrative, the rubric components, they do have an impact on each other.

John also pointed out that the evaluation instrument addressed specific data about which teachers could identify and make connections, and that it was imperative for teachers to pursue identifying specific measures to envelop the strategies outlined in the rubric:

I guess the data doesn't have much value without the details, which would be the rubric part. We have lots of data, statistics thrown at us from the state, and without clarifying information, it doesn't really have a lot of information. Right now, for example, data from the state has come out about top to bottom, not only from the state, but Mackinaw Center. They're very different, but they're different because the data they looked at was different. So, those clarifying points, you could have data that might be a little bit lower, but then without the evaluation information, maybe you are one of the top teachers based on your strategies, your interaction with students, so why is there a mismatch? We'd need more information to look at it. So, the data alone, the data is valuable, most definitely… I think, in past history, even at the beginning in the late ‘80s, early ‘90s, yikes, that's a long time ago, when MEAP was first coming out, people's attitude, “Well, this is never gonna stay around”; well, it's got a different name, it's still here.
The data is one way that teachers building schools are being evaluated in the big picture. There is a need for us to make sure we are looking at both sides of that, not only the data, but the observation. Yep, on a day-to-day basis, does it appear that the evaluation components teacher is strong in, but if students are not performing, then we have to figure out what that disconnect is, most definitely. But I think the attitude part has changed where people have in the past, as I said, the data was kinda like, “Whatever, that doesn't mean anything,” I do believe more and more people are seeing that their individual evaluation is the data.

John continued to note the introduction of the instrument at a global level to the district, but avowed that growth transpired in teachers improving their craft in the pre- and post-observation discourse that followed observations. John was not convinced that the presentations and initial professional development training could be credited for the growth that the district experienced. He believed that he was teaching and modeling with the instructors in a collaborative way as a result of the changes in the process:

We did some general information sessions, but I think probably the more the time spent with teachers individually, in pre-observation and post-observation is the component I feel that helped. Because in a large group, I don't think the sharing, trying to talk and explain, did as much as spending time with people individually. I guess that time of helping people understand, getting people to understand the rubric, what it means, and then maybe the modeling, that next step of modeling the process was the most helpful.

John was also cognizant that the process introduced a unique sense of discomfort for some staff members at the inception of the data being included in the instrument, but he
believed that the majority of instructors became more comfortable with evaluation as observations and discourse relative to instruction increased in frequency:

There still are people who I know, for lack of another word to use, go into panic mode when there's a formal observation or you walk in just because you hear something, you want to go see what's going on so you walk by, go into the classroom. There are people who are truly, you know, I said, *panic* because, “Why are you here? What do you need? What did I do wrong? What did I...” That's one example, but I do not think people have as much of a divide anymore in regards to anxiety associated with being evaluated…we talk about it so much more.

John seemed to assess that the frequency of evaluation mandated by PA 102 (2011) provided the unanticipated consequence of alleviating some of the fear that educators in the district initially incurred. John expressed that simply because administrators and teachers were compelled by statute to spend more time with each other, the development of more pronounced relationships predicated on working towards improved educational outcomes occurred—Evaluation is not as ominous; it's not as abstract, maybe as it was at one time, because there is more sharing.

**Former administrator Anne.** Anne achieved her bachelor’s degree in elementary education and earned her master’s degree as a reading specialist at Central Michigan University in 1984. Anne was interviewed at her home.

Anne began working with the district in 1978 as an elementary teacher and was later promoted to Title I director, where she served for 20 years. She simultaneously assumed the role of curriculum director for 12 of those years. In 2006, Anne assumed the role of principal
of two elementary buildings for four years before closing her last year with the district as curriculum/Title I director in 2011.

Anne became a trainer for Zimco Corporation, which marketed materials and trained districts across numerous states in the process of teacher evaluation and facilitated the automation of the method for many districts that were attempting to conform to PA102 (2011). As a trainer, part of her responsibilities included the actual identification of what data each district would want to incorporate into their individualized and identified structure of educator evaluation.

Anne credited the ability of the district to rapidly move through a process and quickly integrate a tool for staff to be individually measured in reference to student achievement data because the district had already spent a tremendous amount of time and resources implementing the Danielson (1996) portion of the tool just a few years prior. Anne conveyed that the training specific to accepting that the Danielson model was what quality instruction looked like did not need to be rehashed because those discussion had already transpired prior to 2010. The collaboration that developed prior to PA 102 (2011) proved integral to the inclusion of student data being connected individually to instructors in the district:

That year, I would say there was not a whole lot of training on whether or not we all agreed good teaching looked the same. It was more in implementing the tool. And I think, in the district, we implemented our tool really well, because I was there when we implemented Danielson. In the first few years, we did it all on paper. But I know, with my staff, and some of the other principals' staffs, we had the teachers evaluate themselves. We walked through it with the teachers. We referred to it with the teachers throughout the school year. When it came to the point where we were
actually implementing it as our evaluation tool, there should not have been any surprises for teachers, which is crucial.

Anne explained that student data were looked at, but from a view that was a much broader spectrum than what the integration of student achievement data in 2010 would reveal. Attention to specific measurable student outcomes transcended what individual teachers or the district as a whole had previously done. Much more attention began to be associated with individualized teacher performance as a whole. The analysis, relative to how each teacher was teaching, became more engrained:

When we used to look at our MEAP scores, let's say Math scores, if we were low in a certain skill, say estimating, teachers would try to implement more estimating things, but it never really was linked back to how the teacher's performance was. As far as an evaluation tool, we were trying to raise our kids test scores, but never really looked at the actual... Other than, “Everybody needs to be working on estimation,” as opposed to, “How are you teaching estimation?”

The illustration provided by Anne was coupled with concentration that an objective measure needed beyond achievement data. The Danielson rubric (1996) provided that platform because not all data are quantifiable and easily recorded in an automation system or on paper and pencil. Anne contended that she experienced more difficulty in her capacity as curriculum director and working with staff K-12. She shared that the secondary staff as a whole had a more challenging time capturing relevant qualitative data, because the students are not with the educators as much at that stage of educational development. Elementary teachers spend a tremendous amount of time with a smaller number of students:
I always found as curriculum director working with the secondary staff, and this is not a slam to them, it just was much harder for them to capture the information, as opposed to teachers who were working with a smaller number of students… not all data can be captured on paper and pencil or even the computer. There still has to be an objective way of observing students and observing teachers.

Anne credited her familiarity through experience with working on data for allowing her to rely on her teaching background to evoke reflection from her staff toward what the data illustrated by student performance should be. Anne seemed to welcome the changes inherent with PA 102 [2011], as the district had already developed common assessments and the law allowed her to leverage meaningful conversations as to what the assessments would or could reveal. Anne concentrated on the notion that there were multiple measures engrained in the InFORMED rubric that the district had been using prior to the launch of PA 102 (2011). The ensuing comparisons to how teacher data was compared relative to their colleagues proved a natural course of the process:

We created assessments for our district, not the standardized test, but our common assessments with the little kids especially, we looked at changing how we were evaluating our kids and seeing if we were getting any differences and if we were getting consistent data. Was one person standing out far above the others? If so, what was that person doing? Is it because we're not implementing the assessment correctly? Was it because we are implementing correctly and they were a superstar teacher? Was it because of this, that, and the other thing? And my staff, I had a great staff that could sit and have these conversations with each other, with the exception of a few. You always have...people like that.
However, Anne consistently said that the inspection of the data forced conversations independent of whether the teacher wanted to acknowledge the data or not. Anne felt that she now had the ability to retort to teachers who had previously wanted to neglect acknowledgement of student growth or lack thereof. In addition to the common assessments that Anne recalled, she directed credit to the district introducing NWEA into the tool at its inception, as the staff had been trained in this program and had already garnered experience with analyzing the NWEA data piece:

I think, as far as the InFORMED part, once the data could actually be linked to the kids, I don't know how the staff couldn’t look at it. I don't know this for sure, but I would assume that NWEA would have had a huge help with that, because they could consistently see the data without waiting for it. Like with the state test and things like that, you have to wait for the return, and plus, you're always evaluating a different group of kids. Unless you link it back to the individual kids, I think it's hard to judge, 'cause then you're comparing apples to oranges again. Whereas with the NWEA data and our common assessments, I think it was eye-opening. I know as an administrator, when I would do an evaluation, and I'd say, “Look at your data. What do you have to say?” it's kind of hard for them to say, “Well, I don't agree with it.”

Anne referenced the accountability component of the process numerous times in her discussion. Yet Anne contended that the emphasis was not just on proficiency but growth from the onset which was assisted in her estimation through NWEA being involved at the onset. Anne felt that she was a key leader in the construction of the evaluation instrument and that the design commenced with attention being empathetically directed towards
equitable student distribution and identification of growth to be fair to all teachers regardless of the student ability levels with which they were presented:

I think the biggest thing with the union was the fear that I'm a special education teacher, I'm never going to have good data or I always get the bad kids or the low kids… I would always tell them, “We're not looking at, 'Are they performing at grade level?' We are looking at their growth…are they growing as much as those special education kids are?” That's what the evaluation tool's intent is. We were not looking to see that everybody was proficient, we were looking for the growth…we did tweaking and the growth as we did InFORMED. We started looking at the growth more than... that's when it got more difficult to do… “how do you show growth?” But that was always my intent. The tool was not, “Are we all performing at grade level?” but, “Are the high kids growing as well as the low kids?”

Anne attempted to assert that NWEA was a key component to lending legitimacy to the tool but that no single piece of data can inclusively cover all aspects of what needed to be considered to ascertain how effectively a teacher was delivering instruction. That said, Anne was determined to drive her point that procurement of student data absent the inclusion of the dialogue and common understanding of what good teaching looks like is a meaningless endeavor:

Any tool works, I think it's something that has to be almost as much teacher-driven as administrator-driven. And I'm not even so sure that the kind of data that you collect is as important as everybody knowing what good teaching looks like.

**Former board member, Clark.** Clark served the district on the board of education from 2010 until 2014. He was elected by the board to serve as president for three years and
completed his last year of his term as vice-president. Clark has a background in human resources and employee development in the private sector. He is a graduate of the district who has pursued higher education in the field of business. Clark is no longer associated with the district or education. Clark was interviewed at the researcher’s home.

Clark recalled the direction that the board had given the superintendent during 2010. He was convinced that the board was very clear at the time that, as soon as the law passed connecting student achievement results to individualized teachers, collectively the board expected prompt action on behalf of administrator(s) to comply with the law:

The board's decision was to be proactive in developing a system and a tool that would allow us to comply with the act in question. There was a significant emphasis on student achievement data into the tool that was developed.

Clark contended that the main goal was simply to be compliant with the law and that the push toward effective teacher evaluation was one of the biggest issues he encountered as a board member early on:

Teacher evaluation was one of the first big items that came before the board after my election…it was a very big item, it took up a great deal of time very early in my tenure with the board. I remember it because it was more significant than I thought it should be. Coming from a few years in a management role in the private sector, the fact that results weren't being measured and employees being held accountable for those results just baffled me at the time, and how significant an emphasis on student data was from the perspective of my fellow board members and even some people involved in this school district at the time surprised me.
On an individual level, yes, I was surprised. I mean, at the end of the day, teachers are, in my opinion, some of the most valuable and under-recognized assets in our country, and to know that some of the popular conceptions about teachers and their profession being *cushy*, it was clear that those conclusions could be drawn by the fact that teachers really weren't being held accountable for how their students performed. Now I understand, the teacher isn't the only factor that influences a student's performance data. I believe that there are physical limitations, in some cases. I believe that there are emotional limitations, in some cases. And I believe that family plays a significant role in how a student develops. But at the same time, to not hold the teacher accountable for how their students are performing would be like evaluating a CEO by how well he dressed as opposed to the return on his investment for the year, or how he managed capital for the company.

Clark drew on the aforementioned analogy to illustrate that he was not familiar with how educator evaluation transpired prior to his election. Clark contended that there was no resistance from him in regard to ensuring that a system was quickly integrated. He acknowledged that the changes being asked marked a watershed moment as they were complex by virtue of legislative accompaniment, and to what it compelled districts to conform. Clark felt the epicenter of the legislative mandate was an easy concept to digest:

I think anytime a regulatory body changes regulations in today's day and age, it's overly complex. It has to be for legal standards and things of that nature, I understand that. But at its core, if you can boil down to the base concepts behind all of the double-speaking, fancy lawyer-talk, they're very simple concepts that I believe a lot
of people, at least in my experience, hold true. You should be evaluated based on how your students are performing, is a very basic concept to a lot of people.

Clark expressed his pride in the accomplishment of the district and the trust that was delegated to the administrative team in executing the directives distributed by the board:

One of the things that I remember being most proud of about being on the board that first year was that the vast majority of the members of the board understood that managing a school district from a board's perspective requires a certain amount of trust in your superintendent and in your administrators. One of the most significant jobs of the board is to hire a superintendent, and why on earth would you hire a superintendent that you didn't trust to help lead his or her team on projects like this?

So to get more to the point and answer your question, I believe that the board's role was more of a leadership role in saying, “This is the direction we'd like you to head in.” The administrative team, including the superintendent, took that direction and developed the tool with a lot of personal time involved as well.

Clark recalled that the entire administrative team was focused on making sure that the evaluation process was constructed with fidelity and in an effort to construct support to alleviate the associated criticism that inevitably develops with change. Clark suggested that the support commandeered by the administrative team was essential to decompressing the resistance that surfaced with identifying teachers not performing to an acceptable level:

I think it was something the administrative team thoroughly embraced. I know that it was something that the board thoroughly embraced. I think for something like that to be successful, it starts at the top. Everyone has to embrace the concept that, “Hey, this is going to be a good positive change.” The more resistance you get, the more
complicated the issue becomes, because people were looking for an excuse to say, “No, this isn't the right thing to do.” Could it be that some of those people were people that turn out to be underperformers? Absolutely. But you're always going to get resistance to some of these things. The key here was the board setting the direction and letting the administrators develop the tool so that they could absolutely embrace it.

In regard to development of the tool, Clark felt that the administrators had trust that the board was supporting them as educational leaders to engage in the objective of ensuring that a system that measured student achievement was integrated. By ensuring that the system would be in place, Clark felt this led to improved perception of the district in the community and across the state:

When you can get a team of leaders invested in a project, they're leaders for a reason. Leaders then go out and get the people that report to them invested, and it trickles down. I will also say that as a corollary to their performance data improvement, I believe this project helped drive a significant push in driving a better public opinion of the school district. I think this was a pinnacle moment for this school when the board decided to embrace this and develop that tool, and the way that they went about it, I think, had significant impact long term, not just on how teachers perform their job, but how the public sees the district.

Clark appeared to be very pleased with the outcomes in regard to the impact that he felt the evaluation process contributed to the district and its improved performance. Clark referenced the negative atmosphere that existed prior to getting elected to the board and felt
that the impetus behind his desire to serve was improving the district and that the evaluation initiative was critical to his personal objective:

One of the reasons I decided to put my name into the running to sit on the board was because of a lot of the negative perception that had surrounded the district. I mean, this is the school that I graduated from. I was very proud of the fact that I had graduated from this school. But I also saw a lot of room for growth in this school. Would I have ever bet that the school would see this much growth in this amount of time as far as effectiveness is concerned? It's not a bet I would have made. Did I hope? Absolutely. Did I believe it could happen? Absolutely, but it was a cautious optimism.

The relative contribution of the evaluation process at the district toward achieving measured improved student data results was essential, according to Clark. The confidence of his assessment is reflected in his statement regarding the significance that PA 102 (2011) and the ensuing implications of the law:

Changes in teacher evaluation played a significant role. Again, that was one of the most measurable changes in terms of process and procedure at the school, and it would be foolish to discount the idea that it had a significant impact on the performance results as we see them today.

Clark was quick to point out that security measures were critical to the process of evaluation, as the district is situated in a small town, and the perception that some evaluation processes could be viewed through a lens of vindictiveness was a legitimate concern. Clark felt that a combination of quantitative and qualitative data needed to be considered, and he recalled the board conveying sentiments of that nature:
Finding a balance between the things that you can measure that make a successful teacher and the things you can't measure quantitatively that make a successful teacher. I think putting too much emphasis one way or the other defeats the purpose of an effective evaluation tool. Again, too much qualitative data and you leave too much room for interpretation; too much quantitative data and you're taking out the human element… I think the board's role was, at the time again, not necessarily a hands on development role. It was more of a, “This is what we would like to see accomplished,” and putting the faith in the very intelligent people that we had in the administrative roles in the school to accomplish that goal… I'm not going to credit the evaluation tool 100% with the changes, but I think this was the starting point that led to the significant turnaround that we've seen; again, not only in performance data but in public opinion as well.

While Clark could not draw a precise quantifiable number to associate the process of teacher evaluation contributing to measurable district success, he felt the relationship was very strong. He offered the following analysis: If it was measurable, I would be willing to bet that there is some correlation between the two points. The cause-effect relationship.

**Current board member, Buck.** Buck was elected to the board in 2011. He served two years as treasurer and the balance of his term as board president. Buck is an alumnus of the district. He pursued higher education in business and finance and is employed as a human resources director in the private sector of manufacturing industry. Buck sees his role as a board member as representing the public and ensuring adherence to the laws of the State of Michigan. Buck has made a strong effort to become trained in the arena of school board
service, as he has pursued and achieved credentialing from Michigan Association of School Boards (MASB). Buck was interviewed at the researcher’s home.

In reference to evaluation and the implications associated with PA 102 (2011), the superintendent was tasked with developing an instrument to facilitate the demands of PA 102 (2011), according to Buck. He recalled that the tool had been crafted prior to his arrival on the board and before the law was amended in 2011:

I remember that we worked with a marketing company to try and get that tool out, even presented it to the state, I believe it was overlooked. I thought it was a good tool, but the state went a different direction with the tool that they kind of got behind. But we did develop a tool prior to my board service. I think that we were just a little too ahead of our time on that one.

Buck believed that in the seven years he has served the board, the focus toward student achievement data is much more than when he began his first term and drives discussion, whereas when he initially was introduced to the board the emphasis was still on curriculum:

It's certainly a focus whenever we talk about academics in the board room. It's usually focused on achievement and growth as opposed to... It used to be focused on curriculum and graduation rate, things of that nature. And we still discuss curriculum, but usually as it relates to achievement and growth.

Buck conceded that the understanding what is associated with the law related to educator evaluation is complex for a number of reasons and that, despite his efforts, he is not certain that he fully has come to grasp the implications of these changes. Buck carefully annotated that the individual dynamics inherent with understanding the data on an individual
and collective basis can be difficult to fully decipher. He also relayed that the design of the tool was complete prior to his service but that he has played a role along with the board in ensuring that it moved forward. He strongly believed that the district should be able to utilize its own tool as opposed to being mandated to pick a state recommended option.

Buck felt that that evaluation changes has had an impact on improvement, but he downplayed the significance. Buck acknowledged that the district analyzes data much more as a result of the tool, but that the climate and attention to school improvement as a whole marks the biggest contributing factors to the positive achievement experienced in the district:

I don't think the evaluation tool is the largest part of our success. If I were to question where we would be today minus the evaluation tool? I'd say, we'd probably be in a similar place. I think because of the school improvement atmosphere here and the teamwork on the school improvement teams, down to the building teams and the department teams…I think they have really made the difference. Now, are they using student achievement and growth data? I'm sure they are. I'm sure they're using it more than they did before, but I think because of the team that's been built to handle these issues, I can't say that we wouldn't end up in the same place we are now without the achievement tools, the growth tools, the evaluation tools that we've used.

Technology director, Kristina. Kristina has served the district in the technology department since 2002. She was promoted to the role of technology director in 2007. Kristina has received certification and credentialing in the arena of educational technology through Michigan School Business Officials (MSBO). Kristina oversees upwards of 1,800 devices, two employees, and all network communications on behalf of the district. Kristina was interviewed at her office.
Kristina identified the charge that she was given in respect to constructing interface systems that would corroborate student achievement data to specific educators:

The direction, for me, from the board was actually to work with Zimco hand-in-hand, and our student information system to extract the data and upload it into STAGES. Also, help with the training and assist with anything needed.

Kristina offered that it was a challenging process, as the district had also recently just made a major adjustment in changing student information systems. Particularly taxing for Kristina was that the teachers were all being assessed differently at the state level and tracking the students was a trying circumstance for Kristina:

It was differentiated because kindergarten through second grade, the configuration of the district at that time was all in one building, and they were being evaluated differently and the data was deemed tied, which was unique because, at the time, the state assessment was the MEAP and that only began in the third grade. So all those teachers, K-2, were feeding into the MEAP, so we were trying to track all that data as well. I didn't even think about all that, just how different and challenging that was K-2.

Kristina noted another major obstacle, that the district had not had a consistent grading component until 2007. Improvement has been tremendous according to how the collection of data has progressed from the perspective of Kristina. Concern over the authenticity of the data appears to be something about which Kristina felt strongly:

Collecting data—it has improved tremendously. I do have concern that the teachers may be worried more about how the data looks versus how they actually teach now. Teachers are so concerned with every precise little point that might nick their
evaluation a little bit, and it's always referred to, “This is going to affect my evaluation.” They are concerned with how every student may impact their evaluation instead of teaching them, and whether the student is actually learning or not.

Kristina invoked a sense of sympathy for the alteration of how she perceives many teachers are looking at education now. She implied a degree of deflation and defeat has overcome many teachers with whom she works:

It's almost like they forgot about what they became a teacher for, and now they're too worried and concerned about how everything will reflect on their evaluation. Instead of the joy and enthusiasm that I'm sure they got into teaching because they love to teach and mold the minds of young learners, but somehow I think the state is making that more impossible for them.

Kristina conveyed that more autonomy should be returned to the educators in the classroom. She offered the sentiments that a simple pre- and post-test should be sufficient for ascertaining whether a student was successful in a particular course or not in conjunction with state assessments:

Whoever is mandating all the evaluation information? I think my perspective is if they were to use a legitimate pre- and post-test for evaluation purposes, and the state testing for evaluation purposes, and leave the NWEA for them just to determine growth for their own meaning of what they're teaching, whether the kids are getting things or not. I feel that they would be teaching more meaningfully.

Kristina ranked the challenges of extracting data from each source and cited again the issues inherent with not initially being familiar with a new grading system. Kristina acknowledged that mining any of the information at the onset was trying. Local grades
became the easiest piece to collect followed by NWEA; state testing data was the biggest hurdle to cross, and it was eventually dropped from the evaluation tool due to the complexity of aligning state assessments to staff:

Everything was hard when we first started, because we didn't have a true legitimate easy way to extract data from anything to put it in there to begin with. But then NWEA has actually been the easiest of all of them to do, besides local grades… Zimco also changed how they wanted the data input… But with STAGES trying to get the data put in to STAGES nice and neatly and the InFORMED part of it, there was some, as you say, growing pains there along the way, because things would change each year to try to get a little bit better, a little more efficient about getting data into it.

Kristina recollected that since 2017 the evaluation instrument has changed, some years drastically in regard to what data was incorporated, and in other years there were no changes. Kristina felt that it was an accurate assessment that she had to learn a new language and become more acclimated to the instructional delivery side of her job. She felt that this led to comparisons to her identified industry of technology in the educational field. The comparison that educators were having to move quicker and adapt at a faster pace than that typically experienced, which Kristina assured is modus operandi in the technology field:

Everything changes on a daily basis. There's new updates, new upgrades, everything, which has a trickle effect on everything else…in understanding the evaluation process more going through all this has helped me. I have a better understanding than I did when I first started.
The sense of empathy that Kristina conveyed showed in the willingness with which she personally confronts change. Kristina felt that when she is confident with harnessing the change and sees the benefit, she is welcoming to change. However, Kristina commented:

But if it's something that I feel that might slow the district or harm the district, then I would like to put the brakes on and see how it works out for someone else first…We had no choice…it was, this is what we're doing as a district, so jump on and go…the board gave the directive.

Kristina did not feel nervous about making the changes and viewed them as just a challenge in her job. Her job admittedly changed dramatically with the inception of the evaluation piece. When Kristina was asked to reflect on the percentage of her job that was dedicated to working with student data prior to 2010 and what it is now, she responded with the following:

Student data is probably, if you're talking just for evaluation purposes, probably about 30% of my job a year as opposed to almost none before…not for just the student side of it, but for the administration side of it as well.

When asked if she has received more personnel resources in her department, she reported that she had not incurred more support. She has worked at training the teachers to upload their own data into the system and, according to Kristina, this has resulted in an easier process that occurs on an annual basis:

Now teachers upload the PDF as the evidence and they put in the percentages, they type in the percentages…teachers also have to fill out the spreadsheet for the pre- and post-test. At the beginning when we were doing it through STAGES, it was kind of like a spider web, the way they were trying to do everything. Zimco wanted to have
the student ID numbers in there and then you would upload the information to the student ID numbers, and then you'd also have to have another file that tied it to the teacher IDs…if you didn't get that in properly, it really skewed the data…trying to get through that process, because InFORMED wasn't really easy to tie.

In reference to ascertaining if the district could credit the evaluation process to the success it has incurred, Kristina was not convinced that it has had a major impact, partially due to her admitted idealistic perception of why teachers enter the field: “There are certain teachers that I think have stepped up to try to meet and exceed expectations through this. But I think there are some others that kind of are just resistant to change.”

Kristina believed that initially teachers would manipulate data, in regard to testing students with NWEA, up to five times until a student scored where the teacher was comfortable. Kristina offered the following as she felt that unethical practices were germinated as a result of the increased emphasis on test scores:

It concerned me that Johnny scored below what he scored on his very first test and that a teacher would keep retesting Johnny until he got a better score….it would come back to me that, “Oh, well, Johnny is just having a bad day,” or, “Johnny just wants to do better.” But yet sometimes they would test a kid up to five times. That's where I went to administration and had a fit…that has stopped.

Kristina felt that if the stress associated with the testing could be alleviated, the process could encounter improvement. She was not able to offer how this would happen, but she felt strongly that the concern over test scores evoked so much stress on teachers that it stunted growth:
I think I would like to see us keep trying to get better. Maybe looking for different options to get better. Maybe if, and this is a big if, if we could find some way to alleviate the stressfulness that the teachers feel with this evaluation tool, then maybe we could really use it more advantageously than we are…One teacher was in tears just the other day because of trying to get everything done to make sure they have data to upload…They get stressed out, and in our everyday jobs, we get stressed out…to have this added to them, I think it takes away from their teaching…I'm not saying we can do away with it either, though. We have to have some type of evaluation tool. So as long as we can always continue to be thoughtful, that's not a good term, but trying to find better ways and not changing it every year though at the same time… if we can do the impossible, in the perfect world it'd be great…I do see a lot of stress on some of the teachers, that they're just so wrapped up and worried about the evaluation tool.

Kristina pointed to the stress as a continued barrier to educational advancement and genuinely appeared to express concern for the pressure that was exerted on the teachers in the district as a result of the emphasis that was mandated and associated with student achievement data being directly linked to the teachers.

Former teacher, Mikey. Mikey is a former female elementary teacher, who taught for 27 years. She was employed in the district from 2000 to 2014, following prior teaching experience in Alaska. At the time of this study, Mikey worked for the district and earned the substitute teacher daily rate for a total of seven weeks during the year. Mikey administered NWEA testing and state assessments at the elementary level. Mikey received her bachelor’s
degree in elementary education from the University of Alaska in 2001 and the master’s
degree in the art of teaching in the summer of 2008.

In recalling experiences related to evaluation, she contended that at the beginning of
her tenure with the district, the anxiety associated with evaluation was not as intense as it
would eventually become:

I remember at the beginning it was personal evaluations from my supervisor and goal
setting, but I don't remember a lot of paperwork. I'm sure there was some, but it
doesn't stand out… I was comfortable with it, and I got to know my supervisors pretty
well. They came in and out of my room as they pleased… it was very comfortable
when it was not structured. When I knew they were coming, I always got nervous
about it, but they always seemed to go pretty well. I didn't see a great helpfulness in
all paperwork. It made me feel like maybe I wasn't doing enough right. And so, this
new evaluation had to come in and I saw it as a distraction that I didn't want to be
bothered with, to be honest with you. It was state mandated, but the district
paperwork associated with it made it less personal. There was stiff paperwork that
took away from the interpersonal process.

Mikey conceded that eventually she was forced to look at student data more keenly
but that initially it was a considerable hindrance to her. Mikey also admitted that she knew it
was not the right attitude to have, but that it was just a reality. Mikey considered the change
at first to be a personal affront that she was not doing her job despite scoring highly on her
evaluations:
It took a while for me to realize that it was just a generic thing that they were doing, it wasn't...directed at me...my personality is to always want to do it right, and so not being real sure about that caused me anxiety.

Mikey pointed to a lack of confidence in working through the technological parts associated with the evaluation and the paperwork data that had to be pored over and reviewed along with being openly scrutinized. According to Mikey, the confinement in which she had to teach seemed to stymie her capacity to be creative within the curriculum:

It took away a lot of the freedom that I had, and I did not like that...we did, of course, we sat down, and I remember talking about various things, like on whatever grade level that we had, that these things were not being met, the kids weren't doing well in this issue or this issue...so that was good and we could brainstorm a little bit, but it brought a lot of fear among teachers, and that was a lot of the conversation instead of, “How can we use this good tool?” Or “This tool?” I won't say if it's good or not...I didn't like it. From what I can gather now and observe, no longer teaching in the district, from what I gather, from what I'm seeing now, I like better now what's happening. I feel like it's gotten figured out a little bit. People are more comfortable with it...people are more comfortable and kids that are coming out of college now are so much more technologically prepared. That was another issue that was going on, we were learning technology at the same time that we were learning this new evaluation tool, and it just seemed like too much.

Mikey was candid in sharing that part of the impetus behind her retirement decision was associated with the evaluation process and issues she confronted with her colleagues.

The reason, according to Mikey, was that her teammate had a narrow vision and would not
want to engage in cross-curricular learning opportunities for students for fear that if her subject was not being taught, it would adversely impact her evaluation:

I'm a person who would like to use all of my resources in a way that I see fit; most of the time that was easily done. I worked with people who were like-minded, but when you work with someone, and I do not mean to be insulting here to anyone, but when your personalities are so different, I want to be able to do this much in a unit and some people have a very narrow view of what their responsibilities are…that was hard for me. I couldn't come up with a new “Let's try this,” because that was... It just would not happen, and I would attribute it to my teammate being hung up on her evaluation…”this is what I am responsible for and that is what I am going to do, this is what you are responsible for and that is what you should do” would be the responses I would encounter. It discouraged collaboration among elementary teammates, in this case third grade.

Mikey did feel that the integration of NWEA was a data valuable component, as much as identifying goals for the children as serving as a benchmark to measure educator effectiveness. Her reservations regarding the equity associated with tests administered by the state were expressed in no uncertain terms as she felt that the impacts of state testing are so intense:

I do like the NWEA. I like it because the kids have... They know what they're looking at. They, because they see it three times a year, they have a feeling for, “This is what I'm going to be expected to know in science or math or whatever.” And so, I think it makes it that much more important to the kids, which is a valuable tool if you're trying to be evaluated. If they're into it, they like it. They like knowing what the
numbers are. They like knowing that they're moving up. Students who are really bright are very aware of that number. Kids who are less aware know that they're still moving up. And so, I see them all trying a little bit harder, so, that I like. And I think for teachers, it's a reaffirming thing as well. Most of the kids, they're gonna do better each time. And if they're not, there's a reason. Let's look into that. So I do like the NWEA. I don't care for the MEAP because it's a one hit; all the stakes are on that, riding on that, and that's just not fair to anybody. Not the students, not the teachers, not the district… you can't help but want to teach to that test. You can't help it, especially if your evaluation is tied to it.

When inquiring as to what Mikey’s perception of her own effectiveness on her evaluations would be, she insisted that they “had to be highly effective.” A review of Mikey’s evaluations found that she was annually rated as highly effective.

**Former teacher and current curriculum coordinator/content coach, Krista.**

Krista was first associated with the district in 1976, as it was her site for student-teaching. She was hired in 1978 and taught upper elementary through high school during the course of her 38-year teaching career with the district. The last two years prior to retirement, Krista assumed the role of content coach, where she worked with teachers who were improving their own instruction. Krista earned her general education degree in 1976 and her master’s degree as a reading specialist from Central Michigan University. Krista was contracted back to the district as content coach and to also oversee the curriculum development and integration within the district. Krista was awarded the American Legion Teacher of the Year Award in 2013. A review of her final five evaluation cycles indicated that she was considered highly effective on all of her evaluations and had an average rating of 3.74 out of
4 points. The lowest component of her evaluation were the student performance results, which averaged at 3.27 points (highly effective) of 4 points.

Krista recalled the dramatic changes that ensued since she was initially introduced to the district. She recalled and offered a personal anecdote relative to how she earned tenure, and how the process became much more involved:

I spent some time on lay-off, and when I came back, I didn't have tenure yet. In the spring, I went to my principal and said, “Don't you need to evaluate me?” He said, “Oh, no. You're a tenured teacher, I don't have to evaluate you”. I said, “No” They had already sent the tenure to the board, so I got tenure by default…I never even got evaluated that year because…I don't recall how often you had to get evaluated, but you didn't have to get evaluated every year. But because I was supposedly tenure, I didn't even get evaluated my second year either…it got serious when we received the Danielson evaluation around 2006. Prior to this, I had one evaluation…it was just scripted…everything I said, it was worthless…that was my evaluation.

Krista acknowledged that the pendulum had shifted from where she initially experienced evaluation to where it was currently. Krista did not feel that the process was exactly where it needed to be at this time, and that evaluation is not the only driving force behind reviewing student achievement:

I think we've gotten much better at analyzing data. I don't think there was much of an effort to look at data...we looked at data and things like that, but it never got to the point that it drove your instruction, and I think that's where we're heading... It's still all so new…things keep changing and whatnot…We did use data to drive our
instruction, but I think it's just getting more intense than it used to, but I think it's more school improvement process that drives looking at data than the evaluation.

In retrospect, Krista reflected on the Danielson component compared to the InFOMED data portion. She believed that there was too much associated with evaluation, and she reiterated that she is grateful being retired as a teacher that she does not worry about it in the same way. In relation to the data, she has concerns with the validity due to the way some staff approached the assessment process with students:

I understood part of the data, but again, I think some of it's skewed. I think grades, and we've had this conversation before, grades as part of the data became very skewed. I think even NWEA, when you have people who say, “Take them into test in the fall and tell the kids, ‘don't worry about this test,’” and then get all intense about it at the end when growth is measured…I think somewhat it's skewed, and so therefore some of it's irrelevant.

Teacher input through school improvement process has been considered, according to Krista. She did believe that there was no choice when the district directed that Danielson (1996) was being implemented. Krista reiterated that the data from students is reviewed by teachers, and discussion and action as a result ensues:

The school improvement team at the high School was pretty much the one who drove the decision to not have NWEA included. But there are consequences to everything…I think, especially if I look at the elementary school in particular, I think you have a lot of teacher-driven decisions. Again, I'm not so sure that it's driven by the evaluation tool, but more by a group of people that just plain want it to be better…those kind of people that just want the schools to be the best they can be. I
think that's driving more improvement internally than the evaluation process....some people just plain want us to be better, regardless.

Krista was concerned about the universally inclusive terms inherent in the evaluation tool and expressed that they served as barriers to teachers as they were intimidating to the tool’s size and scope. She expressed that this can become detrimental to the highest achieving teachers who strive for excellence:

I think we need to have a focus. We have a shotgun thing where we've got how many sections are there in that thing, and there's no way…some of the verbiage in there, your best teachers, as they read that, are going to say, “I can never be this, because it says all students will.” Or some of the words that are in Danielson. So as you look at it, your better teachers are the people that are most reflective, are going to look at it and say, “There's too many of these inclusive words, all students, every student, those kinds of things. Nobody is ever there.”

Former teacher Brooke. Brooke began her teaching career in a charter school and was hired by the district in 2010. She served as a second and third grade teacher in the district until leaving in the fall of 2016 to assume the role of assistant principal in a neighboring district. Brooke was a member of the school improvement team throughout her tenure in the district. Brooke earned her bachelor’s degree in 2007 and her master’s degree in educational leadership in 2013 from Saginaw Valley State University. Brooke averaged 93% of all possible student performance points on her evaluation over a 5-year cycle. She was considered highly effective on all of her evaluations. Brooke was new to the district the year that the student achievement data was to be integrated. She recalled that her experience with the evaluation process was relatively smooth:
Evaluation went smoothly for me but I was a newer teacher, so then I was evaluated or observed, I guess, you could say, more ...it was simple for me. The evaluations went well...I had the same evaluator the whole time.

Brooke saw little change in regard to her perspective associated with the evaluation assessments during her tenure with the district:

I think some teachers, well me, I am all about the data and I want the numbers to look good, but then other teachers are just like it's one test and it's not a big deal. It's about what we seen in the classroom that matters more than anything else, so I think it's probably a combination of both the teachers that use the data... it did help in the classroom, and then other teachers just kind of foster relationships with their kids better instead of just evaluating the data...I think seeing numbers helps. When I would help teachers put in their data onto the stages, them [teachers] seeing the numbers helped...I had to help people figure out their percentages and the data because some teachers had a hard time looking at it, I had a handful of teachers I helped every year.

Brooke did not feel that she could offer as much insight as she had hoped, but she expressed a major consequence of the evaluation tool and the shortage of teachers that she perceived currently existed:

You have to look at the effectiveness evaluation, and if it's even doing anything. Right now, with the way we have a teacher shortage going, the evaluation process is going to be harder to do because you don't want to get rid of teachers, because you can't replace them in a sense. You really can't. So, and then being...what are you going to be able to reload it with?
Brooke expressed that her frustration stems from her awareness that the state still had not provided direction in regard to evaluation, and she was frustrated with the resources that have been expended attempting to ascertain the best processes in various districts.

**Former teacher Fran.** Fran earned her undergraduate degree in 1977 and master’s degree in education in 1986 from Central Michigan University, and she taught numerous elementary grades during her career. Fran spent the largest portion of her career teaching students in the district’s gifted and talented program. From 2010–2013, Fran taught fourth grade math to students of all ability levels.

Following her retirement, Fran returned to the district for six weeks annually to proctor NWEA and state testing to students. A review of Fran’s final evaluations indicated that she moved from being considered effective on her evaluation in 2011 to being considered highly effective at the time of her retirement.

Recalling her experience with teacher evaluation at the onset of her career in 1977, Fran felt that her evaluations were primarily narrative in nature and simply done to satisfy the district’s requirement of two evaluations each year, per the collectively bargained agreement:

> When I first started teaching, principal did two evaluations per year on non-tenured people and you were non-tenured for two years. You get tenure at the end of the second year but you had to do two visitations. He did and then he met with the new teachers individually, talked to us all, and gave us a formal write-up. Most of it was essay paragraph form…As time progressed, there were more multiple choice, more checklists per se. I know for new teachers, the tenured law changed and you had to go four years, I believe, and so those... Being an experienced teacher, if your evaluations were acceptable, then you were only checked, I know, formally once a year.
Occasionally, some years it changed. Some years, you had to do two or three visits and one evaluation though...principals came to visit you two or three times, but you only got the formal evaluation written after the last visit. Each visit had to be at least 30 minutes.

Fran conveyed that changes in her opinion were needed in regard to the alterations in the evaluation process after the integration of PA 102 (2011) and the introduction of student data into the tool. Ironically, in Fran’s estimation she did not look at data much differently than she had in the past. Fran credited this as a result of her exposure teaching gifted and talented students, where data were necessary to keep children in the program:

I'd had academically talented children in my class. So we were looking at data all the time for those students to make sure where they were, that they were above the data and I'd always watch the data. Having had the kinds of kids that were put in my room, we ended up needing to have extra data on them. I really didn't notice a whole lot as far as data goes, because we had extra tests and things to get the kids into my program. They had to test into that program.

The changes that Fran noticed as relevant were when the district changed from not allowing teachers in the room when the NWEA testing was proctored to allowing the teachers in the room with the proctors. According to Fran, not having the teachers in the classroom when the students engaged in testing opened the door for an environment that was not conducive to testing for elementary students:

When the testing happened. We were not allowed in the room, so you couldn't control what was done. You had no control over what kind of a test setting they were in. It was run by a paraprofessional that didn’t have the same influence over the students as
their teacher. Once we could be in and control classroom management, so that you didn't have kids all over the place and could just keep it a good testing situation, it makes a difference to have a consistent environment.

Establishing an ambience conducive to a stable testing environment was not the only challenge with the introduction of different student achievement data to consider. Fran conveyed that there was great anxiety associated with the evaluation process as a whole from her perspective on a personal level and that the complex nature of the changes in the evaluation process heightened her insecurities:

It was a pretty complex change as far as all of the little things that were being looked for, the wording was sometimes... I can't remember. Elementary people don't use all the formal words and things that are being shown in these evaluations. And just to make sure we could wrap our heads around some of the, exactly where the wording was, and what the percentages were, and where everything sat. The evaluation process always threw me a little bit...just someone coming in with a computer or a clipboard or whatever they were using, absolutely set me off. Just a little paranoid, but... I've never had any reason to be paranoid. That's just my own insecurities.

In spite of the trepidation that Fran encountered with the evaluation process as a whole, she found that her last few years of teaching where achievement results and data were included in the evaluation process, led to areas where she could identify specific points to improve. Fran felt that the specific information provided gave her direction to improve her craft. Further, Fran identified that after the initial shock of having to change and adapt through another system that her anxiety associated with evaluation began to recede:
As far as data that I could get back at what I needed to do the last couple evals that I had, that worked because there were certain things that gave me direction to improve. I was like, this one was effective instead of highly effective. What do I need to do to get me up to the next category and to keep going up because... Let's face it, no one's going to be perfect all down the side. So you're just constantly trying to improve… I think once you've been through the process the first time or two, it got much, much better. It also got better for the administrators doing the evals. That first year was not easy for anybody to try to do those and get the results back to us in a timely manner. There was a lot more writing and things that administrators had to do to get it in.

Fran conveyed during the course of the conversation that she did not have a hand in developing the evaluation tool utilized at the end of her career but that she informally worked through its integration, as she shared that there was a lot of stress associated with the new process for many of her colleagues. However, Fran also contended that the short duration that was remaining in her career gave her the perspective that she just needed to ‘deal with’ the process until she chose to retire:

I think when you get to the end of your career, as some of us were, it's like, “Okay, deal with it, we're out of here, anyway, shortly.” It wasn't as stressful as I'm sure it could've been right at the beginning.

Fran believed that in order for the process to improve, evaluations should be more focused on objective and identifiable measures. The processes and effectiveness of those that Fran incurred in her career were altered considerably in her estimation depending on the administrator who was conducting the evaluation. Fran suggested that the literary
competence of the administrator typically played a significant role in the evaluation processes that she experienced:

Evaluation needs to be more objective and less subjective. That was always my biggest thing is it was so subjective, depending on who your principal was. Some of the principals were very talented and were writers, and evaluations were fantastic, and other ones that did it because they had to do it, and that was not their strong suit, had a hard time. Their writing skills did not make the teacher look as good or as bad as they should have…if you had somebody else looking, well, you've got two evaluations, one that somebody could write and one that somebody couldn't write. Which one are you going to look at for hiring somebody else? That's hard for people to understand. When it's all written narratives, it's hard for administrators to do that kind of thing depending on their writing ability.

Fran thought that the introduction of reviewing lessons on video in the same manner that athletes review game film will also be part of the process moving into the future, but she is not convinced she is comfortable with videotaping everything in the classroom. However, Fran acknowledged that this would be a measure of safety and accountability at the same time. Having a complete record of what a teacher does is the way that Fran sees evaluation moving:

There is going to be more video, more audio, cameras in the classroom, but it's going to accountability that way. We watch kids on buses, you watch kids in halls, you have to see kids in classrooms, so you know everything that's going on. The police have to have cameras on body cams. Not that it's going to be a body cam for a teacher, but I
can see a camera running in a classroom for everybody's, not just evaluation, but for safety and protection for students and staff in this day and age…I hate to say that.

**Former teacher Julia.** Julia served the district in this study for 35 years as an elementary teacher. Julia attended Grand Valley State College and earned her degree in elementary education in 1977. From 2010–2013 Julia taught fourth grade language arts. Outside of volunteering, Julia is no longer associated with the district. Julia’s last three evaluations were reviewed and she was considered effective in 2011 and 2012. In 2013, Julia was recognized as highly effective. In all three cycles of evaluation, the student performance data were the lowest component of the evaluation; however, each year there was considerable improvement and all cycles were registered as “effective” in the student data component.

When she began in 1977, Julia recalled the process of her evaluation as a relatively simplistic process that was condensed to one page and that the process evolved tremendously throughout the course of her career, with particular emphasis placed on the changes that transpired at the end of her career:

At the beginning of my career, teachers would be evaluated every year, twice a year. Basically, the principal would come in and observe your class for an hour or so. The principal would then come back and talk with you about what he saw and what he thought were your strengths and weaknesses. Basically that was it, and then you could refute anything that you didn't agree with. It was part of our contract… it seemed like it changed a lot at the end, the last year or so. It had more input as far as student achievement and it wasn't just based on what they saw in that one-hour time
slot. It was based on test scores, it was based on everything a teacher did…not just the one-hour dog and pony show.

Julia contended that by the time that she retired, the evaluation process yielded a much more accurate depiction of what a teacher was contributing as a whole. Julia acknowledged the initial challenges associated with the alterations to the evaluation process and the incorporation of student achievement data, as it was difficult to understand what the information meant. The collaboration that ensued among teachers in reviewing data, however, proved to be a benefit, according to Julia:

I think it helped me understand data better. It was a learning tool for me, too. It was a learning tool…I would say at the beginning, it was challenging. But a good thing about that was that it made you talk with other teachers so that you could make sure that you were understanding it the way it should have been understood. I was lucky enough that I was team teaching at that time so that I had someone else that we could sit down and discuss, and that was nice.

Julia contended that at the time the evaluation tool was being altered and with the process evolving so rapidly, she did not recognize the effect that the integration of student achievement data would have on the district. Upon reflection, Julia felt that the role of the evaluation process was a significant contributor to the success that the district has experienced. However, she was hesitant at the time that the changes were integrated to think that the evaluation tool would result in improved performance:

In regards [sic] to the tool being useful or not when all the changes happened, the jury wasn't out yet on that. I don't think I realized how important it would be, but I really think it benefited greatly…I think it helped educate all involved what we were
looking for, because I don't think everybody used data to drive their instruction or whatever. I think it forced them to, which they may have balked at first, but I think just the fact that we are showing achievement proved that it was good… I think that there was resistance at first. I think pretty much, “Oh no, here's another thing.” But I think in the long run, it's proven to be a valuable tool for us.

Julia expressed that her primary concern was that the equitable nature of evaluating is a challenge when considering the distribution of students’ ability but, even more, the volume of students a teacher has assigned to them. Julia referred to her personal experiences to express that the size of a class is something that needs to be considered in the evaluation process, particularly when the result is a relative comparison to other teachers:

I know that some teachers have like 23 students in a class and others may have 35 and 36. I know for a fact you can be the best teacher on the face of the earth, but the more bodies you get into a room affects how effective you can be. So I think that's kind of hard to, “Oh, this person does a wonderful job,” and they have 23 students. Where another teacher may have students that aren't achieving…well, there's 35 kids in that class. How can you meet all the needs?

Julia felt that evaluations should be an ongoing process inclusive of a lot of information to reach a conclusion of effectiveness. Julia contended that the most effective processes that she experienced involved frequent walk-through observation by principals:

I think the most effective way to evaluate a teacher is to go into their classroom and not be announced that you're coming in. Just walking in. I liked that a principal just came in and stayed for a few minutes. First, you'd think, “What's he doing?” But a principal can come in and kind of absorb what was going on, and I liked that. Because
to me, well, it is nice when a principal is visible. But I think you really see a teacher at their best when they're not doing the dog and pony show, but when they're actually just...day-to-day teaching. What they're doing every day. So I think that's effective.

**Teacher leader Melania.** Melania is a special education instructor who has served in the district since 1997. She worked with students in grades kindergarten through second. Melania is the district track and cross country coach in conjunction with serving as a member of the negotiating team on behalf of the teacher bargaining group. Melania’s bachelor’s degree is from Michigan State University in 1994 and her 2000 master’s degree was from Saginaw Valley State University. Melania earned a rating of “effective” on all of her evaluations from 2011–2016. In all of 2016, her student performance results were rated as minimally effective, but it is worth noting that she served the lowest ability levels in the district. A search of her previous evaluations indicated that the principal did not enter a student performance rating in 2012 and 2013.

Recalling her experiences when first being introduced to the evaluation process, she remembered that the process was less involved and much more informal from her perspective than what would come later:

Evaluation was much less strict than it is now. I remember being evaluated maybe two times a year. It seemed like when I first started before I had tenure, maybe it was two to three. It was just I believe, handwritten, and then we'd meet after and talk about the lesson; definitely wasn't as formal as it is now. We didn't have a formal large document with all the domains and areas that we use now. Once I was tenured, I believe it was every couple of years that we were evaluated. There wasn't a yearly
evaluation like now; it feels like every two years it seemed, maybe it was four at some point, it wasn't very often.

The process evolved, according to Melania, and she recognized drastic differences in regard to what the district was introducing to teachers by implementing the Danielson model of evaluation. The initiation of a new model seemed to be characterized by Melania as overwhelming, due to the size and scope of the tool:

I remember reading through it. I remember going through it, and it was huge, and it was kind of overwhelming as to all those... It's quite wordy, I guess it was, there was a lot of verbiage in there that needs to be understood. I think it definitely holds us more accountable in all the areas. I'm not sure all of it's pertinent and useful information when I read through it. There's so much to it, and I'm not sure if I get marked down on something. I've questioned my principal what do I need to do different to improve that area and... he's given me examples of things... Sometimes, some of the requirements on the evaluation I find are... I'm not sure how useful they are, but I guess just because it's so much, but it has been cut back now… some of the language in it was confusing. I didn't really understand what they wanted us to do or what they were asking us to do.

Apparently, Melania welcomed the reduction in the scope of components that were measured in the document. She was not willing to acknowledge that the system was easy to use in regard to providing a map to effective instruction, but she was adamant that narrowing some of the components allowed her to focus on more specific areas of her skill set:

I don't want to say user-friendly, but at least you can read through the evaluation now and understand it a little more. Before, I didn't go really into detail because there was
so much. If I was marked down in an area, then I'd look at that area, and I did question the principal before. What didn't I do here? What do I need to do to fix that or improve that? But he would give me an answer…we have to constantly evaluate if it's appropriate as things evolve and change in education. You have to make sure your tool is evolving as well, which I think the district has done. Where you've gone from that very long, lengthy document to…it's more streamlined now and focused on maybe more specific components.

The introduction of student outcome data into the instrument also provided Melania insight into analyzing results differently than she had in the past. Melania felt that the inquiry and attention to student achievement data has increased considerably from what the focus had been previous to 2010:

The importance wasn't placed on the data prior to this tool being used. Obviously, statewide, nationwide, data is very important now. Before, I'd look at my specific kids and you look at their growth for the year. Now I look at the scores of all of the kids because when you look at your evaluation at the end, you look to see where your kids are overall to see what kinda progress they've made. There's definitely a much larger focus on the use of data and what it means and how to improve it at this point.

Melania found the student data component of the evaluation process easier to utilize and decipher. However, she was very apprehensive about the accuracy of the MEAP testing that was integrated into the evaluation process at a time when teachers were not entering their own data into the system:

The data is pretty cut and dry, it's graphed out and you can see, I'll put the NWEA scores in and the process is familiar now, we've done that for a long time. The pre-
and post-test is pretty easy to see and use… I don't know how much I trust that MEAP was 100% accurate. Actually, I don't trust that it was, because I remember one year, and I can think of a student I had where, with teaching special ed, most of my kids are not proficient in all areas or even many areas they're not proficient in and this one year, I had several kindergarten students and the one, when I looked at her data, it was wrong because it had her as proficient in writing and she was a student that absolutely was not proficient. I looked at several kids and the data was not put in correctly and, at that point, it wasn't when teachers were putting the data in.

Melania contended that people lost faith in the process when the state testing was involved in the calculation of the student achievement component of the instrument, due to the difficult task of tracking the data from students to individual teachers. Melania recognized the challenges associated with this, as she taught kindergarten students who would not encounter state testing for three more years when the children would take the MEAP for the first time in 3rd grade: “I think a lot of people were finding that, if you can't trust the numbers, it makes it harder. It was definitely more than one student that year.”

Melania cited that the district has been responsive to making changes in the process and the tool to ensure that it is an accurate reflection of a teacher’s data. However, she tempered her endorsement by citing the challenges infused in the summation of ascertaining what is considered adequate growth for students who cope with learning disabilities:

I know that teachers have been offered input into the tool, and when suggestions and complaints have been raised, administration has responded. Administration has asked for insight. I have assisted new teachers in understanding the tool because it is confusing…My question always is the special education aspect of it, because most of
my kids are not going to make a full year's growth or more. If a student is a
cognitively impaired kid, they don't make a full year's growth often. They might one
year. I've had kids make a full year's growth or even more. But then other times, they
were lucky if they make six months growth in a year, it just depends how low they
are.

Melania believed that more attention must be given to the special education
component of evaluation. She provided an illustration of a student with whom she is
currently working that highlights her point that some cognitively impaired (CI) students
develop in unique ways that she feels must be considered:

I have one this year, he wasn't reading. I mean, very low, kindergarten level. And
we're half way through the year, and he's jumped to later first grade, and he's a CI
kid…then, he might level off and not make progress for a while. Those are the tough
ones when you're being evaluated on these students…there is a reason they are
special ed. They have disabilities, and it's always a concern.

**Teacher leader Fred.** Fred is a secondary science and social studies teacher who has
practiced his craft exclusively at the district since 1997. Prior to that, Fred finished high
school in the district and later graduated from Ferris State University in 1999 with a degree in
education. He earned a master’s degree from Saginaw Valley State University in 2000 and
an educational specialist degree in curriculum in 2011 from Grand Valley State University.
Fred served as the chief negotiator for the union membership from 2006–2016 and currently
serves on the negotiating team for the union. In 2013 and 2014, the district rated Fred as
highly effective on his overall evaluation and on his student performance results. It is worth
noting that in 2015, Fred saw 76% of his students meet their target growth rate on NWEA;
92% of his students met their goals on pre- and post-assessments, and 99% of his students earned a D- or higher. Despite Fred’s growth on student achievement, he earned an overall ranking of effective for 2015.

Fred recalled that when hired in 1997, only a few points were looked at from a pedagogical perspective. Fred was transparent in his perspective in regard to the evolution of the evaluation process and the impact it had in his 20 years of experience. Fred made it clear that he wanted to give an accurate representation about how he viewed the process and the way that it has made him feel. In reference to the introduction of Fred’s exposure to the evaluation process, he marveled at the expanse in growth that has transpired:

When we started, it was basically, you got a check mark satisfactory or unsatisfactory. And there were maybe, and I might be stretching it, maybe eight or 10 points that were looked at, and there was usually a written summary at the bottom that said something flattering about what the administrator thought about you and your class…a brief narrative just about your overall function in the school; that was about it. I would say, from maybe 97 to 2005 or 2006-ish, when we started doing the Charlotte Danielson rubric, and we've been doing that and modifying the rubric since then. I'd say, I'm not sure exactly on the time period.

Fred made it a priority to highlight that the progression was a collaborative effort among the union leadership and district administration. Fred was able to establish the procedure the district typically uses on an annual basis to understand the delivery of the evaluation process as it has evolved every year:

It's been basically a cooperative effort between administration and the OEA, the bargaining unit. Usually the negotiating team, if problems occur or if an idea comes
up with one side or the other, the two would get together and go over what the new plan is to try to alleviate any problems we may have had the previous year…Our August meeting is when we do this. We have a meeting before the first of the year, so during the August meeting, if there were any problems either with data or any other parts of the evaluation, we address it at that point.

Fred offered insight into the state’s mandate that four levels of effectiveness must be incorporated into evaluation instrument. Fred seemed frustrated that the district did not have discretion over this component of the tool. He offered an alternative to the ineffective, minimally effective, effective and highly effective components of the tool:

I'm not sure about the four levels. To me, highly proficient, proficient, minimally and not effective, maybe a three… There's no middle. You're either right at the bottom of proficient or at the top of minimally. And if it were a proficient, I don't know if average or satisfactory and then not proficient, I don't know if that would be an easier evaluation tool, because a lot of times there's quite a bit of discussion between teachers and administrators as to, “Are you in this category or this category?” I don't know. It might just make it easier as a tool.

In regard to overall improvement of the process, Fred felt there was a lot of positive impact. For example, he highlighted that he definitely personally looks at teaching and learning data significantly differently than he previously had as a result of the connection to his measured effectiveness being tied to the data:

I would say before 2010, I can't really recall looking at student data either… There were test scores, standardized test scores, but as far as how many passed or failed classes, that really was not looked at until suddenly, if a student failed your class, it
affected the teacher's evaluation. Initially I believe the student had to have a C minus in your class in order for you to have a satisfactory evaluation…I think there were a lot of C-minuses out there.

When asked for clarification, Fred was able to point out that he felt that teachers as a whole might have presented students who typically would have received Ds to now be considered effective, the student would be perceived in their class as a C minus:

I am saying that there were a lot of C minuses at that point, because the student was receiving credit for getting a D, but the teacher wasn't receiving credit on their evaluation for the student having passed the class, unless they had a 70% C minus.

Fred provided insight into the evolution of this phenomenon by highlighting the progression of education having the onus on the student to do the work to now the teacher is responsible for the student achievement. Fred contended that being forced to scrutinize data and its implications created a definite shift in mindset individually and collectively. He also offers his own personal experiences in regards to an exhibit:

I would say as the school and also as myself, yeah. If the kid didn't do the work, I mean, it's just changing of education, if the kid didn't do the work, they got an E…”You should've worked harder,” and they weren't all the intervention steps we now have, and there wasn't the help that we have in the learning center, where there is two certified teachers offering assistance in small settings. It was, I guess, “You should've tried harder.”

I can remember one biology class, it seems like out of 20 kids, 9 of 'em failed. And it didn't even faze me, and I was a rather new teacher, within my first 10 years, and now if you have that, it's changed the atmosphere, but I would say for the
positive. I think there's more effort to help the students in the classrooms and also with the out-of-class, after-school tutoring, after-school workshops, lunch workshops.

Fred was optimistic and appeared proud that he was part of a shifting mindset; however, Fred was clear that there were many hurdles to get to the juncture that the district is currently stationed, from his vantage point:

I think when we first went to this rubric, it was a new thing, we were learning continuously, there were issues coming up; there were challenges, and then when we added data, which I believe was a state thing, wasn't it? When we had to add data, there definitely has been some growth in the student data realm, as far as technology to be able to figure out with the data, because I know originally I had sophomores, and if I would've had any juniors, that would've taken, at that time the ACT, it was probably because they had failed biology, but I think a third of my data was based on those nine kids that were retaking a 10th grade class and how they did on the ACT. Which, guess what? Wasn't real good because of the type of students they were.

And that was the first year, and initially it was supposed to follow from one year to the next, it was supposed to be easier to track kids through. At the initial onset of the data in the evaluation, they were going to be able to, even at fourth grade, give credit to the first, second, and third grade teacher for how well the student did, and it just didn't work out; and there were students that were on class lists that you never had in class, and it was all technology; and we worked through that, and kinda the last year before we started doing the data as teachers, I believe everybody got 100% for their data. I know I did. Because it was too difficult...everyone was given 100% on the state testing because it was too difficult to track. The ambition
was there to tie students to state testing all the way through, but technology
couldn’t support it and the testing kept changing so dramatically.

Fred seemed pleasantly surprised about the impact that evaluation has had on the
district. Of particular distinction, Fred was taken by how it has actually been positive impact.
This marked a shift for Fred as something that he identifies as an alteration of a systemic
approach to the delivery of educational philosophy that encompasses the district:

I wouldn't have thought it would've had a big impact at that time when it initially
came out. It may have also almost appeared to staff that it was an administrative way
to get you with looking at the student achievement, but I think the staff has adapted,
and I don't think they're playing the system, necessarily. However, I think they
realize if a kid doesn't get it, you need to not just give up. The kid isn't going to get it,
you need to try a different approach. You need to call home, set up a lunch workshop,
after-school workshop, something like that. So initially, I would have said no, but I
think it definitely has improved, or the perspective of the teachers has changed. The
data is important, we need to look at that. We need to work to improve student
achievement and I think we're in the process with the kids. I just still don't think the
kids think the tests are important. That's a culture thing that we need to still work on.

The improved resilient approach that educators must exert to working with students
who have not placed an emphasis on their education is a reciprocating venture, according to
Fred. His sentiments reflected that the students are not in the same place as educators in
regard to the investment on the tests. Fred thought that student input could be interesting in
an evaluation, and that it could be looked at. Fred was not a proponent of parents or other
teachers evaluating teachers. Particularly other educators, as he felt self-image could play
too big of an impact: “People have distorted perceptions of themselves. They use that to compare to others and I just don't think that would be a good avenue, seeing what students perceive could be interesting.”

In regard to sustainability of the tool, Fred expressed that the district has evolved well and should be wary to change whenever the state changes. Reflecting on how the district has worked through the qualitative component of the tool, he feels that reducing it down to smaller chunks was a positive:

I think initially there were 65 steps or thereabouts in the Charlotte Danielson and we've now kind of bundled those into groups. I think that was a good step. I have no problem with the tool that we have now. I don't think it's a bad tool, and I think, for just consistency, to be able to look one year to the next to the next, it would be best not to do major changes in the face of the state, which every couple of years wants to have a new test for a new measurement, and you can't compare it back more than two or three years, because it was a different test and a different measurement device that they had. If we can get the data tool figured out at the state level, whether it's an NWEA type of test, or something like that, so that everybody is apples-to-apples, I think we need to do something like that. Sticking with the tool we have now…I think this would do it. I see no reason to change it.

Fred did proceed to highlight the challenges associated when an evaluator changes. He noted that some parts in his experience could be construed as subjective because of his own personal experience when his principal changed:

I think it's rather subjective. I know that my evaluation tends to change. With the Charlotte Danielson that we're using, it has changed dramatically as I've gone from
one administrator to another administrator to another one. Over the last three administrators, my evaluation has gone down, down, and down, but I believe my administrators would tell me that I'm becoming a better teacher. But compared to what initially was...considered looked as highly effective, just effective, minimally or not, and I'm just pulling numbers off my cuff here but I'm going to say I was close to 30 highly effective, two administrators ago. That number dropped down into the high teens, or low 20s with the next administrator, and now it's right around 10 to 14. I haven't dropped into the minimally but it's just that there's been a drop back.

But we're learning more about it, and I know administration's going and actually learning more about how to use the tool. But there has been a definite change. So if someone were to look at that...I have changed as a teacher to the positive as I am looking at data a lot more than I had in the past.

**Teacher leader Bob.** Bob is a highly qualified secondary teacher in the areas of business and social studies. He has been with the district since 1998 when he graduated from Northern Michigan University with a bachelor’s degree in business education. Bob also earned a master’s degree from Northern Michigan University in educational administration in 2005. Bob has served on the negotiating team on behalf of the union since 2001 and currently maintains that position. Bob serves as one of the two teachers servicing the learning center at the district. The two teachers work with Grades 7 through 12 to ensure that interventions are properly offered. Bob explained succinctly:

In the learning center, kids come to me to get help on their assignments. I run a report every Monday morning to find out who is failing. I try to get as many kids passing, as possible. I assign after-school workshop when possible or when needed,
as well as lunch academic workshop, when needed. At these times students get additional interventions and support from myself and other teachers in a variety of subjects. The learning center started in 2011, and is funded through categoricals.

In reviewing the process of evaluation since entering the district, Bob cited how the process has become more structured. He was rated as either “effective” or “highly effective” on his evaluations from 2011–2016. In 2015, Bob earned 3.75 out of 4 points for his overall evaluation score. It is worth noting that in 2015, Bob earned 98% of all possible points on the Danielson portion of the instrument and garnered 3.04 out of 4 points possible on the student performance data. Bob’s data scores are the conglomeration of all students at the high school:

The evaluation is now a lot more thorough. I know, my first four years, you had to get evaluated every year…twice a year. The principal would come in or assistant principal would come in for one class period and that was it; you didn't hear much feedback at all. The principal would just take notes…In 2003 or so the district brought in Danielson. Then it seemed like the principal cared more about it.

He(principal) said, “Try new things.” So he'd give you a comment, you could ask him for help and then it seemed like they were trying to help you more at that time. Before then, it was more like, “Hey, this is what I gotta do…lets get it done.”

Bob believed that it was powerful for him, as he took a college course on the Danielson tool and also worked with the district at the inception of the instrument. From his role in the bargaining unit, Bob realized that the district was moving toward a defined tool for evaluation in 2002 or so and he consulted the Michigan Education Association Uniserv director (UD), who endorsed the tool. Bob claimed that the district had to negotiate to use
their own tool at the time and the district had language related to the evaluation process taken out. At the time, Bob felt that anxiety was high among staff—that administrators wanted to target specific teachers for removal:

Dave Sabedra was here; he was our Uniserv director. I remember telling Dave, I says, “We need to have an evaluation tool and I studied the Danielson tool. The administrators are out there trying to help you improve, they're not out to get you.” That's what the feeling was, that all these administrators are out to get you. Dave felt also that if we had to have a tool, it was good…that was the feeling. Danielson was presented as a tool to help you get improved, we'll help you get training, we're going to help you do this, this and this to help you get better. And if you want to try something new, we are here to give you feedback. You started seeing that now, lately, is, “Here, this is when I'm coming in, what do you want me to do, what do you want me to look for? What can I help with?” that is a nice approach. There are pre- and post-evaluation meetings with the principal or evaluator.

Bob looked ahead 14 years from when the process began and noted that the tool continues to change every summer. Bob pointed to an example of how initially in 2010, the district set proficiency of passing to be 70% but that staff felt that if students could pass with 60%, then teachers should be afforded the same consideration:

Teachers were looking at going from 60% as proficient to 70%. We said, “You know what? Passing is 60%.” So this is part of the input that the staff had. Why would you need to be above that to be proficient? Teachers have had input but the state only gives you four options on what tool to use anyways. My gut feeling tells me that the elementary teachers have more influence over the evaluation than high school
teachers. High school teachers will bitch if something goes bad, but they don’t care about the data as much if it doesn’t… Evaluation does have an impact, but you learn to play the game. Your evaluation... I don't play the game, because I am evaluated as the total high school average, and I don’t assign grades. I know that teachers don’t count tests anymore where kids did bad. They retest them over and over. They reteach it or some might just not count them at all. I am being honest…they modify tests for students to be successful based at their ability sometimes, too. Maybe it is differentiation, but it is a lot different than it used to be for sure. Kids get modified tests when they aren’t even special education.

In respect to the possible inclusion of peer, parent, or student input being incorporated into the tool, Bob does not feel that there is merit in any of those possibilities. He is adamant that the teacher perspective is critical for the continued improvement of the process. Without the communication between the evaluator and the teacher evaluated there would be a breakdown that is difficult to overcome. Bob provided an example of this in regard to his own professional experience and perspective on the evolution of evaluation:

The main thing is that you have to have teacher input. You have to work through the pre- and post-observation meetings. The teacher has to let the person know what's going on, the person who's evaluating you. The evaluator has to give feedback…if they didn't see something... One time, I got evaluated, and a girl was sick, and she had her head down, and she was sick, you could tell she was sick. She didn't want to go to the office. So I went up very quietly, just went up to her, nonchalantly, and said, ‘Hey Mary, I see you're not feeling well. Is there anything I can do for you?’ ‘Nope.’ I said, ‘Okay I'm going to let you stay here.’ So then I went back, and went about with the
rest of the lesson. I came back and said, 'Hey, are you alright?' The administrator never saw it and he put that in my evaluation. Kid had her head down the whole time, teacher never went up to him. 'No, I went up to that person twice. You never saw it.' So in the post-observation, I tell him, 'Well, you just never saw it then, but it did happen. Then I have ownership in the process to make sure it is accurate, which is what we want as teachers. We understand that evaluation is a process to make the whole school better.

**Teacher leader Tim.** Tim has been an elementary teacher with the district since 1996 and has taught in Grades 5 through 7. He has earned his bachelor’s degree from Central Michigan University in 1996 and obtained a master’s degree in educational administration, with emphasis on the principalship, from Saginaw Valley State University in 2001. Tim became the chief negotiator for the bargaining team of teachers in 2016. Tim has been considered highly effective upon a review of all of his evaluations from 2011–2016. In 2016, Tim earned a score of 3.81 points out of 4 points possible on his overall evaluation. Tim earned 82% of the possible points on the data portion of his evaluation in 2016.

Tim recalled the process and evolution of evaluations in the district since he was brought into the fold. Initially, Tim felt that evaluation was of relatively little consequence but that it has evolved considerably:

To the best of my memory, I would say evaluations really weren't anything of note. Pretty much, if you knew you were doing the best you could do, administrator would come in from time to time. It was absolutely different in that realm that administrators had more time to just come in, and I want to say hang out, but participate, share with you what was going on. So I think a lot of those visits, I don't think I ever saw as
being an evaluation-type process; it was more or less just almost like a social, ‘What's going on?’ Where now, it seems like there's a lot more emphasis put on every time an evaluator comes in. There's a lot more stress put on educators, because you want to make sure that you do make a good impression, for better or worse… I am not saying we didn’t used to have follow-up meetings, but I just don’t remember them outside of maybe a quick little talk… I don't remember if there was a paper that was signed to say that we had a meeting even... Like I say, I don't remember them ever being anything that really made me feel like there was any stress involved.

In a way, the whole process has become negative because it did seem so much friendlier back then. For example, the principal would routinely have time and would come in… back in those days, we used to have newspapers every Friday that would be delivered. So for language arts, I would always find articles that I thought the kids could handle. I'd make up questions and we'd almost do like a treasure hunt of sorts, where they'd have to try to find the article. The principal would come in and he'd spread out with the newspapers. He'd get down on the floor with the kids. It was, like I say, it seemed different. Now, of course, I was much younger then, so I maybe wasn't aware as much of what was going on, but now it's like every year you see that list come out of when your observation's going to be done and you mark that on your calendar, and you make sure that no matter how you're feeling that day… you've definitely got all your ducks in a row and you're ready to go for that. Now, that can be good as well… depending on the person, but generally I think if you always are trying to do your best, there's not a whole lot of positive that comes out of having to put yourself through that stress.
Tim did express that reflectively, he reviews student data much more closely. Citing that the scrutiny of that data in a skilled educator can have positive implications. Tim is conscious that this has been a progression that has transpired over his career:

I honestly don't remember looking at data a whole lot back in the beginning of my career other than when the need would come out. Now, being that we have different versions, whether it's the M-STEP, NWEA or our own local grades…I think there is a lot more realization of the global picture of a student's academic progress, which I do think is a good thing, the more information you have. It's the same thing with team-teaching. When you teach by yourself, you don't really get anyone else's perspective. So if I have a student who struggles in math, and then I find out from the other teachers that that student has... That they're doing great in other subjects, okay, maybe it's a math thing. It's not necessarily the student overall who just doesn't care or is a bad student. So I think the more data you have, assuming that you know how to use it, is a positive.

Tim was capable of producing an illustration from the day of the interview where multiple measures of student achievement were utilized to identify intervention groups for targeted instruction. This specific information, according to Tim, will help in servicing students at their ability level:

I don't know all the ways that you can use student data, because you don't know what you know until you learn something new. But for example, today, trying to make a list for our new intervention schedule, we've got the DIBELS tests, we've got our NWEA scores, and we have the previous M-STEP scores. When we look at all of that together, based on what we also know of the kids in class…I think, a more
effective way of then divvying up the class lists so that each kid gets put in a spot
where they're going to learn the most; it gives you more of a picture of the student.

In looking and inspecting himself as a teacher by virtue of the whole picture data
provided in the context of student achievement, Tim was less hesitant to make that
connection as quickly because of the varied ability levels of the students he is charged with
developing:

Considering the data as a whole connected to myself, it depends. For example, I have
three different classes. I teach math, and they are distinctly different from each other.
So, while I know that I'm teaching basically the same way to all three groups, some
groups do better. So generally, when I have that, I have to look at the group
separately, and I think that if I only taught one group, I don't know that I would have
that ability to realize that, ‘Okay, this might work for two of the classes, but it's not
really working for this one particular one.’

Tim is reticent to offer a whole scale assessment of the ability to ascertain the
effectiveness of evaluating based on data, but he readily acknowledged that looking at data is
pivotal to assist the students. Tim seemed genuine and sincere that his goal is to educate, but
he acknowledged that there is a considerable fear associated with evaluation. Tim did not feel
that the changes associated with evaluation were that challenging to comprehend, simply
change was hard to handle:

I don't think the changing of evaluation is really all that difficult to understand. It's
simply that it was a change, and because of the weight that evaluations have in terms
of this could probably cost you your job or help you keep it, whichever way you want
to look at it. Any change, when all of a sudden the stakes are put on the table like
that... becomes stressful. But I think, generally, the understanding of evaluations and how they work isn't really all that difficult.

Tim felt that administration has been helpful in regard to the process and how difficult it could have been. Tim acknowledged that the administration and the union have collaborated to make the process work. However, Tim cited the individual administrators he has served with:

I don't know that I've ever had an administrator that I did not feel would listen to a legitimate concern, and I hope that always stays that way. I don't know if everyone feels that way but I definitely do, so that's a positive.

Tim was able to elaborate even further as he acknowledged that globally, the evaluation process may have had an impact on overall school improvement. However, he explained that, due to his intrinsic motivation and expectations for himself, he was not motivated to work harder per se:

Personally, I don't think that the evaluation tool has made a huge difference for me because I've always felt like I've tried to give my best. I'm not saying every single person does. But your fear motivates as well, and I guess if you are scared that, 'Hey, maybe I'm on the bubble.' Then, yup, maybe that does motivate you to try a little bit harder. But we've made a lot of changes since then as well, whether it's consolidating schools, getting new programs.

Technology is another one. Back in those days, go back five years and you had teachers that wouldn't step into a computer lab because we'd only have one. It was hard to get into, and then you didn't really necessarily know how to troubleshoot any issues that came up, and it became more hassle. Well now, with the technology
the way it is, I think more people do use them. I have a tablet cart here, which I would probably say three times a week, we actually get to get online to the math program, where the kids can actually play math games designed by the program to build those skills. If you can make learning fun, especially with today's generation being more tech-savvy than any generation previously, I think that the chance for making gains is just...exponential. I don't think about the tool when I'm with my kids. I end up thinking about the tool more when I'm by myself and I'm looking at it that way. Once I'm with the kids, it's like all about...getting them from A to B.

Tim considered the possibility of utilizing peer, student, or parental reviews. He was opposed to all of these ideas with the caveat that he could see how observing other teachers in a non-punitive forum could prove helpful. In regard to improvement in process, Tim viewed it as a necessity that the teacher perceives the process as beneficial to them if the evaluation protocol is to have the benefit of having merit associated with it. Tim asserted that this can only happen when an evaluator invests a lot of time in the classroom with the teacher:

When you don't see your administrator in a one-on-one setting very often, if one of my previous principals were to say, ‘Hey, while I was in your room, I thought you could have done this better,’ you would immediately feel like, ‘Uh oh, I'm being attacked.’ Even though I know that principal very well, it's like, ‘You're not in my room enough for me to feel like you get everything that I do and why I do it to then start criticizing.’ So, I think, for criticism to work, you have to feel very, very comfortable with the person giving it and trusting that they're not doing it to hurt, but to help…I think that that's a tough relationship to build with the way things are set up.
now, due to mandates on administrator’s time because they have to evaluate every teacher annually, and let’s face it…time is part of that…I don’t envy administrators now.

Tim said in closing that he should point out that there are so many factors to consider when evaluating a teacher’s effectiveness in reference to the countless dynamics that are out of the grasp of educators to influence, as students come from such unique and diverse backgrounds. Tim highlighted that the idea just does not sit well that a teacher is evaluated on student performance when the student is the biggest factor:

I would say the biggest issue that I have with teacher evaluation is when you base it off of student achievement, based on the kids you have from one year to the next, or even one class to the next, can change so drastically, that, again, it's a little off-putting to feel like, ‘Okay, now I'm going to be viewed on, well, a kid who has missed 35 days this year, or a kid who hasn't even turned in a quarter of the homework.’ There are just so many things that we cannot control. In a perfect world, if every kid came in and sat down, ready to go. And we still have the free breakfasts, so no kid should have hunger as an issue, at least in the beginning of the day. But if every kid sat down, ready to go, prepared, and gave their best, I would be open arms to however student achievement falls, because I would know they did their best, I did my best, and at the end of the day, I can hang my hat on that with pride. But considering there are so many factors out of our hands, it's always a concern. Now, again, everyone's in the same boat, so law of averages, it's all going to play out, but that's not a whole lot of consolation if you're a worrier.
**Teacher leader Mikayla.** Mikayla has served in the district since 1987 and has predominantly taught in kindergarten. She has served as the union president since 2014. She earned tenure in teaching at a previous district prior to 1987. Mikayla has been ranked highly effective upon an examination of her previous five evaluations with the exception of 2016, where she was rated as effective. It is ironic that her lowest score happened to be the year where her student performance data was the highest she has experienced. She compiled 89% of all possible student data points in 2016 but was ranked as effective, opposed to the trend of being considered highly effective on evaluations where student achievement data score was not as high.

Mikayla said that when she was first evaluated, it was simply a short narrative that offered minimal feedback. Mikayla contended that improvement has transpired because of current guidance and direction:

> I think there is a lot more direction in the evaluation process now. At that time, it was basically, they came in, they looked. It was what they saw, what they thought. At least now you have protocol that you have to follow, as far as things that you look for…it helps you when you're planning and trying to be a teacher, to know what people are looking for, what they expect. Generally it has helped me but…I’m not sure of everyone. I don’t know how much people look at it because it is a long document.

In consideration of how Mikayla has reviewed data since it has been incorporated into the evaluation process, she offered that she is more aware of the progress of students but to what degree was not clear:
I probably spend more time looking at my data, but it's probably made me a little more aware of how students are performing. But I'm not quite sure it has what made me change a whole lot.

Part of the frustration that Mikayla seemed to share when discussing the evaluation tool was that it is challenging and complex to understand in her opinion. She expressed difficulty in understanding the reporting requirements that the state had and what her evaluation was reported in reference to her effectiveness rating by the district. She claimed to have discussed it with the principal as she thought that her student achievement data in years where it was higher would have resulted in higher effectiveness ratings but that was not always the case. She expressed confusion in regard to what actual components of student achievement were being measured:

I think there's some confusion, and I'm not sure if people did go on and look at issues I expressed. But I said, for me, I'm not really aware of how they come up with the total rating. Even though I know the percentages and everything, because we set-up our own tables. So I'm not sure if the general teaching population would understand how it's determined due to equations and things… I don’t know how much teacher input was really involved in the creation of the tool. I was vice-president at the time but it was done through the negotiation team.

Mikayla did not expect to see the gains that the district has experienced and she did not realize the evaluation instrument would have as big of a role as it has had in the achievement increases as a system in her estimation. However, she was restrained in crediting the evaluation system with any particular degree of growth:
I didn't think the evaluation tool would have this big of a role. I do think it has made a change, because I think people are a lot more aware of what their classes are doing. I'm not quite sure if it's totally tied to the evaluation process. It has definitely been positive change.

Mikayla pointed to the limitations of considering incorporating in peer review due to the potential animosity that could transpire among colleagues. Mikayla also pointed out that kindergarten students would have a challenging time grasping what they were attempting to evaluate their teacher on. Mikayla did not feel that parents spend enough time observing to be poised to offer evaluative insight.

Mikayla believed that the system must remain to be flexible and fluid as the educational landscape continues to shift and evolve. Mikayla contended that the district has an effective system, considering that pre- and post-conferencing whenever something evaluative is transpiring generally transpires:

Ten Themes Emerged from the Perceptions of Stakeholders

Joint ownership of data and process. Former superintendent, Joan, said that the teacher evaluation process was a “team thing,” because the purpose was to “look at efficacy of instructional practices, instructional curriculum, and instructional delivery in a different way. It made educators, the system, the organization take ownership in a way that did not exist before.” A sense of ownership and teamwork developed, as teachers viewed student success related to teacher effectiveness. Administrators established the expectation of that joint ownership. The hope of the administration appeared to be that embracing the change and being excited for it would translate into staff-buy in. Based on the testimonials of the teaching staff, the introduction of student data into the evaluation process and the realization
that it was not going away led to dissecting and digesting data much differently than was
done prior to 2010. Board members acknowledged that this ownership resulted in their
perception of the community’s acceptance that the school genuinely shared the students’ best
interest when individual student-achievement was tie-barred to the success of the school.

**Trust.** The two board members who were interviewed both expressed a confidence
in the administrative team that was charged with inventing and implementing the protocols
and process of a system to rank the teachers. Joan, former superintendent, said that trust was
the contributing factor that brought about the integration of student achievement data and the
teacher evaluation tool with the a large percentage of support among teachers. The
establishment of trust between the teachers and administration, particularly the principal, was
a critical component to the change initiatives associated with teacher evaluation. The goal at
the administrative level was that the teachers would perceive the changes as an improvement
to which administrators, teachers, and students were united. Teachers regularly expressed
that trust and teamwork were critical to changes in their pedagogical processes and practices.
The relationships evolved between principals and teachers were the catalyst for establishing
the framework of trust. These experiences were often forged through walk-through
observations on a regular basis in concert with pre- and post-observation conferencing, where
important dialogue centered on out what good teaching looked like and how it was identified

**Data provided a more complete picture of student.** The data gleaned in the new
teacher evaluation process definitely altered the way that all stakeholders viewed student
achievement, particularly at the individual level. The views expressed by teachers and
administrators translated into more focus on the success of the students. Teacher leader,
Tim, admitted that in years past, mining data was not a focus for determining student achievement, but a perspective offered by test results plus data related to teacher performance offered “a more global picture of a student’s academic progress.” This attention to detail developed into teachers and staff diagnostically looking at how to improve students from one level of proficiency to the next on a more individualized level than in the past. Through this inquiry, goals for students and staff were established.

**Principals focused on discussions of Danielson (1996) model.** The data generated led to use of the research-based Danielson model as the reference guide for what good teaching looked like. Former teacher Julia, recalled,

> At the beginning of my career, teachers would be evaluated every year, twice a year. Basically, the principal would come in and observe your class for an hour or so. The principal would then come back and talk with you about what he saw and what he thought were your strengths and weaknesses. Basically that was it, and then you could refute anything that you didn't agree with.

Former principal John noted that even though it was a challenge, the positive part was that it was part of building relationships, and “even though it was time consuming, it was a very important part of the process.” Teachers expressed that the more time and input a principal was willing to share resulted in increased comfort in the process. The collective objective of principals was to guide discussions of data and observation to a rubric that outlined expectations. Former principal Rod noted, “Conversations centered around how you move up the scale, so to speak. The staff trusted me to tell them what they had to do.” The teachers were “looking for input as to how they could get better,” and “is this what you mean?” would be a common conversation. The administrators quantified the rubric to assign
a numeric value and this assignment of weighted teacher results became the focus on how to move from one level of the rubric to the next. This process led to a picture of effective teaching. Principals mentioned that identifying where teachers were actually performing on the rubric led to challenging discussions. Teachers also expressed that change in the evaluations and scores depended on the administrator assigned to evaluate.

**Responsibilities and roles have evolved.** Changes in evaluation and the associated mandates changed how teachers looked at data and used it to drive their instruction. The changes associated with evaluation definitely altered the amount of time that teachers, administrators, and technology support personnel dedicated to working with student data and evaluation overall. Former administrator Anne noted that, “Once the data could actually be linked to the kids, I don't know how the staff couldn’t look at it.”

Former principal John said,

I think like anything, it's evolved, because I think until teachers and other administrators had a better understanding of that process and utilizing the components of the evaluation, following through on that, talking, sharing, and even making specific recommendations to teachers based on the rubric. Until you do that, it doesn't really mean anything, so I think that's part of that processing.

Principals evaluated more and conducted more guided conversations about instructional delivery and pedagogical process. The technology director, Kristina, saw her role evolve and expand to a great extent; no time was spent working with student data in her original role versus approximately 40% of her time managing the data related to teacher evaluation as a result of the new state mandates. Kristina identified that she was charged with constructing interface systems that would corroborate student achievement data to specific
educators. Changes in role requirements marked a stark contrast from embedded practices in the district. Kristina offered that it was a challenging process, as the district had also recently just made a major adjustment in changing student information systems. A review of personnel files found that teachers were rarely evaluated prior to 2010, but change in the evaluation process has begun to drive so many of the roles in the school district.

Anxiety relative to evaluations increased. Anxiety in relation to evaluation process was considerable for teachers, and the administration clearly expressed that they took an empathetic and conscientious approach to teachers’ insecurities. Principals and the former superintendent all noted that they were very conscious how the changes in evaluation would create a major shift in anxiety. Former principal John noted that there are still people who go into the “panic mode” at the mention of a formal evaluation. The administrators felt that established trust with a considerable portion of the staff helped them to execute these changes rapidly. All of the teachers interviewed mentioned some continuing anxiety related to evaluation but to a lesser degree. Former teacher Fran spoke to “great anxiety associated with the evaluation process as a whole..., and that the complex nature of the changes in the evaluation process heightened her insecurities.”

Teachers expressed concern about the students to whom they were assigned. One teacher claimed that she retired because she felt stress associated in working with a colleague who would not want to collaborate in a cross-curricular approach once students were assigned to a subject and teacher. Another retired teacher commented that she knew she just had to get through a few more years until retirement. Technology and learning how to track the data also compounded the stress many teachers experienced. technology director, Kristina pointed to the stress as a continued barrier to educational advancement.
**Having the Danielson (1996) model in place.** Former superintendent, Joan, noted much training was done at the time the district shifted towards the Danielson framework in 2005 and she conveyed that this helped later in the transition to adjusting to the ramifications of PA102 (2011). Former administrator Anne conveyed that the training specific to accepting that the Danielson model meant that the question of the nature of quality instruction did not need to be rehashed because those discussion had already transpired prior to 2010.

The administration and teachers concurred that the district’s decision to implement the Danielson model and being familiar with it prior to the inclusion of student achievement data allowed for a relatively smooth progression to the state-mandated teacher evaluation system. The five years of working with the Danielson model was a marked shift from previous educator evaluation. Review of staff members’ previous evaluation instruments indicated that feedback prior to 2005 was limited at best.

Teachers had become accustomed to working with four components of the evaluation tool and the introduction of the data component in 2011 led to more conversations and opportunities to work with principals having intentional professional pedagogical discourse on quality instruction. Both principals and teachers felt that the relationships established through this process promoted trust. In addition, a common language appeared to have been instituted through these processes and a portrait of what effective teaching should look like was already being established in the district. To some, there still exists varying levels of understanding in regard to the evaluation instrument across the landscape of the teaching staff. Teachers recognized that reduction in the size of the instrument allowed more focus and directed areas that can be targeted for improvement.
Multiple measures of assessment. One of the persistent themes referred to multiple measures incorporated into the evaluation tool provided a more complete depiction of a student’s achievement. Teachers and administrators insisted that multiple measures would be used for assessing learning and growth. In addition, the validity associated with the instrument was considered to be increased with additional evidence and exhibits. Former administrator Anne concentrated on the notion that there were multiple measures engrained in the InFORMED rubric that the district had been using prior to the launch of PA 102 (2011). The ensuing comparisons to how teacher data was compared relative to their colleagues proved a natural course of the process.

Having the control for local grades allowed teachers autonomy in reference to contributing to the process. NWEA served as a barometer to identify growth, as did common assessments. In addition, the former curriculum director commented that having district-wide common assessments allowed for rapid integration of this component into the tool. The technology director noted that there were too many measures. The challenges associated with the inclusion of the MEAP—tracking it, anticipating when state testing would change—along with the continuous changes of state baseline data proved too trying to include, and the district eventually dropped state testing from the instrument altogether.

Former administrator Anne contended that although NWEA was a key component to lending legitimacy to the tool, no single piece of data can determine exactly how effectively a teacher was delivering instruction. She said further that procurement of student data absent the inclusion of the dialogue and common understanding of what good teaching looks like is a meaningless endeavor:
Any tool works, I think it's something that has to be almost as much teacher-driven as administrator-driven. And I'm not even so sure that the kind of data that you collect is as important as everybody knowing what good teaching looks like.

**Business mentality.** The approach to new legislation about teacher evaluation became a business directive with a priority of identifying effective teachers. Leadership of the board of education and superintendent was conducted in a business environment. Former board of education member Clark recalled that the board’s decision was to be proactive in developing a system and a tool to comply with the new legislation. The consensus that this initiative simply needed to be done provided the mandate to force this change. Having that latitude led the union to adopt the approach of working collaboratively to influence how the tool would function. The administrative team welcomed the inclusion of suggestions and teacher input. The collaborative approach to working the instrument through was useful to accomplish the task of establishing trust.

**Living document.** The autonomy for the district to maintain control over the instrument(s) used and to be fluid and flexible was paramount to the perception of its effectiveness and continued efficacy. Teacher leader Mikala voiced concern that the tool must be a working document that will change along with the teaching climate. Teacher leader, Bob, concurred saying that “the teacher perspective is critical for the continued improvement of the process.” The regular annual review conducted by the administration and union served to address changes that should be considered. All stakeholder groups interviewed expressed a lack of trust in the state in regards to teacher evaluation. Attempting to be responsive to all of the changes in state assessment has proven to be a large task.
Summary

This chapter began with a brief review of federal and statewide teacher evaluation policy documenting unprecedented change since 2010. Multiple themes emerged in the analysis of stakeholders’ perceptions of the efforts in one Michigan school district to comply with a state-mandated teacher evaluation process revealed. Paramount among the themes was that trust needed to be involved in the process of teacher evaluation for improvement to transpire. The district initiated change in relation to teacher evaluation in 2005 and assessments in 2007; thus by 2010, the administrative team was experienced in leading and implementing programs. It appeared that through these initiatives, a common vision of what evaluation should look like, began to take shape. Adding student achievement data with the authority of the law afforded the district the opportunity to converge multiple pieces to complete the assessment of teachers in relation to their peers. The final chapter in this dissertation includes conclusions and implications related to the findings and recommendations for future studies.
Chapter Five: Conclusions, Implications, and Recommendations

Introduction

Public Act 102 (2011) in the State of Michigan required school districts to comply connect teacher evaluation to student achievement data. The implications of this legislation on one northern Michigan school district had not been rigorously evaluated at the commencement of this study, nor were there scholarly studies regarding the personalized impact of the law upon individual educators. This study examined how a single rural school district prepared for the change process and explored the strategies employed to construct, cultivate, and sustain support for the teacher evaluation process. Implications of the new law that has initiated the integration of teacher evaluation and student achievement data are discussed in this chapter, which concludes with recommendations for future studies.

Conclusions Based on Findings

Themes that emerged from analysis of stakeholders’ responses to interview questions provided the foundation of conclusions that can be drawn about the implementation of change in one school district.

Multiple measures of assessment yielded a global picture of student achievement. Persistent themes among stakeholders’ responses indicated that focus on the success of the students and multiple measures incorporated into the evaluation tool provided a more complete depiction of a student’s achievement.

Teachers and administrators insisted that multiple measures would be used for assessing learning and growth. In addition, the validity associated with the instrument was considered to be increased with additional evidence and exhibits. Local decisions limiting the number of variables regarding assessment measures gave teachers control for local
grades. NWEA served as a barometer to identify growth, as did common assessments across the school district, which allowed for rapid integration of this component into the tool.

Teacher leader Tim addressed the benefit of considering multiple measures to realize the global picture of a student's academic progress: “It's the same thing with team-teaching. When you teach by yourself, you don't really get anyone else's perspective.”

Strom (2011) identified students as critical stakeholders in terms defining academic achievement, and Schoeffel, Kuriloff, and Van Steenwyk (2011) found that students believed that grade point average (GPA) was the most critical factor in determining whether success was achieved. From the student’s vantage point, the inclusion of teacher-issued grades appears critical to the quantitative analysis of teacher effectiveness, and should be included. However, the Danielson rubric is dedicated to classroom environment. This study found that teachers’ orientation to student assessment changed in the process of complying with the mandates of Public Act 102 (2011). Data collected in this study indicated that teachers realized the importance of team learning and the benefits of a global picture of a student's academic progress.

**Trust among stakeholders was essential for coping with change.** This qualitative single case study was conducted in the tradition of Choo (2006), who contended that sense-making provides a forum for considering new information and its implications regarding policy and its interaction with individual’s experiences and prior knowledge. All of the administrators interviewed in this study reflected that trust nurtured by dialogue and professional relationships was the central theme in the integration of an evaluation system that incorporated data into the Danielson (1996) framework. Former superintendent Joan
emphasized the concerted effort to tie stakeholders to student data from the beginning to promote a sense of unity of all throughout the process.

Trust was the most prevalent theme referenced by all participants of the study. However, pinpointing and defining the root cause of a diminished sense of trust within the organization prior to the integration of the evaluation system proved challenging. It appeared that the investment of time that principals spent in conversation with teachers about the Danielson (1996) rubric during pre- and post-observation conferencing provided a forum whereby teachers confronted, for the first time in the district’s history, the connection between achievement data and their performance. By addressing pedagogical practices in conjunction with a tangible instrument that identified quality instruction, it appeared that the staff and administration made strides towards bridging areas where trust had receded.

Administrators played an important role in the process of change. Responses of stakeholders reflected an awareness of the importance of the role of administrators in the process of adapting to change in the way teachers were evaluated and the integration of student achievement data. Teacher leader, Bob, noted that the evaluation is a lot more thorough now. Prior to state regulation, teacher evaluation appeared to be a largely informal and perhaps an infrequent process. Teacher leader, Tim, added that to have the benefit of having merit associated with the evaluation, teachers need to perceive the benefit to themselves. Tim asserted that this can only happen when an evaluator invests a lot of time in the classroom with the teacher

In this study, teachers, administrators, and technology support personnel extended the amount of time dedicated to working with student data and evaluation. As an evaluator,
the principal guided discussions of data and observation to a rubric that outlined expectations about instructional delivery and pedagogical process.

**A business-like and collaborative approach helped to offset anxiety.** The approach to new legislation about teacher evaluation focused on identifying effective teachers and student achievement. Leadership of the board of education and district superintendent was conducted in a business environment, recognizing that this initiative was mandated and needed to be done. In a collaborative approach the teacher’s union worked to influence how the evaluation tool would function, and the administrative team welcomed the inclusion of suggestions and teacher input. Former principal Rod recalled that endorsement of the Michigan Education Association (MEA) and local union representatives removed barriers to successful implementation of the teacher evaluation process with all stakeholders engaged. Although teachers expressed a range of concerns, the established collegial relationships and trust among a considerable portion of the staff helped them to execute the required change.

**Practice using the Danielson (1996) model had benefits for success.** Former administrator Anne other administrators, and teachers concurred that the district’s decision to implement the Danielson (1996) model and being familiar with four components of the Danielson evaluation tool prior to the inclusion of student achievement data allowed for a relatively smooth progression to the state-mandated teacher evaluation system. Anne credited the ability of the district to integrate a tool for staff to be individually measured in reference to student achievement data because the district had already spent time and resources implementing the Danielson (1996) portion of the tool just a few years prior. The introduction of the data component in 2011 led to collaborative work between teachers and
principals about quality instruction linked to student achievement data. Both principals and teachers felt that the relationships established through this process promoted trust and a shared understanding of what effective teaching should look like.

**A collaborative process led to effectiveness and continued efficacy.** Collaboration played a vital role in establishing an effective teacher evaluation process. Former teacher, Julia, acknowledged that initial challenges associated with integration of the evaluation process and incorporation of student achievement data was difficult to understand, but collaboration among teachers in reviewing data however proved to be a benefit.

An important theme that emerged from stakeholders’ interview responses was that the teacher evaluation process is a living document, one that will evolve over time. The autonomy of the school to maintain control over the instrument(s) used and to be fluid and flexible was paramount to the perception of its effectiveness and continued efficacy. The regular annual review conducted by the administration and union served to address changes.

**Answers to the Research Questions**

Three questions guided this study. This section discusses the response to those questions that emerged from the data.

**Q 1. How did one school district prepare for the implementation of Public Act 102?** An important part of the preparation of the school district in this study to implement mandates in PA 102 (2011) included their experience in the use of the Danielson Framework (2007) at the time the statute was enacted; however, the ranking of educators and inclusion of student achievement data had not been done. The district being studied was one of the first districts to aggressively integrate the mandates inherent in PA102 (2011).
Teacher evaluation and student achievement data were gathered and analyzed using the modified Danielson rubric (2007) and InFORMED from 201–2016 through STAGES, a web-based system designed to automate staff evaluations. Student achievement data through InFORMED software integrates student growth and achievement data for an individual educator’s evaluation. Former administrator Anne welcomed the changes inherent with PA 102(2011), as the district had already developed common assessments. Anne noted that there were multiple measures engrained in the InFORMED rubric that the district had been using prior to the launch of PA 102 (2011). The law allowed her to leverage meaningful conversations as to what the assessments would or could reveal. The ensuing comparisons to how teacher data were compared relative to their colleagues proved a natural course of the process. This instrument allows users to choose which data pieces to integrate into the evaluation process.

A combination of information sources, termed triangulation, and perspectives gathered from in-depth interviews of stakeholders increased authenticity and accuracy of the data gathered in this case study. In addition, the study yielded information about the impact that the district’s process for implementing change relative to PA 102 (2011) had on determining whether administrative support and approval was sufficient to support the sustained change needed to benefit students.

Learning outcomes were scrutinized to assess if the InFORMED rubric was an appropriate mechanism to define success in the realm of student learning outcomes, as identified by Danielson (1996), and whether stakeholders in the district interpreted the process of teacher evaluation to have played a role in increased achievement as identified by the Michigan Department of Education (MDE).
Q 2. What strategies were employed to construct, and sustain support for the modified teacher evaluation? Strategies to construct and sustain support for the modified teacher evaluation were revealed in the responses of 17 stakeholders who participated in open-ended interviews. Intentionality to comply with state mandates was demonstrated by the business-like approach taken by the board of education. Former board of education member, Clark, expressed that board trust in the administrators supported educational leaders’ efforts to develop a system that integrated teacher evaluation and student achievement. This key strategy established an environment of optimism and expectations to meet the challenges of the task.

Cooperation and collaboration among the stakeholders, including administrators, staff, union representatives, and policy-makers was a positive strategy for building confidence and overcoming anxiety in the midst of change. Former principal, John, highlighted the relationship(s) being cultivated among teachers and administrators as critical. Principals’ strategy to maintain open and continuing dialogue with staff guided discussions of the use of student growth data and a rubric that outlined expectations and a picture of effective teaching.

Q 3. What did the district learn from the modified Danielson rubric and the InFORMED data results rubric? The administration and teachers concurred that the district’s decision to implement the Danielson model and being familiar with it prior to the inclusion of student achievement data allowed for a relatively smooth progression to the state-mandated teacher evaluation system. One of the persistent themes referred to multiple measures incorporated into the evaluation tool that provided a more complete depiction of a student’s achievement and altered the way that all stakeholders viewed student achievement.
Having control for local grades allowed teachers autonomy in their contributions to the process. NWEA served as a barometer to identify growth, as did common assessments. The autonomy for the district to maintain control over the instrument(s) used and to be fluid and flexible was paramount to the perception of its effectiveness and continued efficacy. Former principal John believed that the Danielson integration at its inception began to formulate individualized dialogue and discourse aimed at improved instruction, but it was the quantifying of the tool and the student achievement that made the difference in the process becoming more persuasive.

Input from stakeholders led to ownership of the process, which in turn, led to collaborative activity across all levels at the school district. Board members acknowledged that this ownership resulted in their perception of the community’s acceptance that the school genuinely shared the students’ best interest when individual student-achievement was tied-barred to the success of the school. Union leaders accepted a positive role in the implementation of the state mandates, and the views expressed by teachers and administrators translated into more focus on the success of the students.

The establishment of trust between the teachers and administration, particularly the principal, was a critical component to the change initiatives associated with teacher evaluation. Former principal John considered the Danielson rubric and the dialogue with educators individually and collectively that ensued in the review of the student achievement portion of the instrument led to key factors in the district’s growth and academic acknowledgement that the district experienced. Teachers regularly expressed that trust and teamwork were critical to changes in their pedagogical processes and practices. The introduction of student data into the evaluation process led to dissecting and digesting data
much differently than was done prior to 2010. Changes in evaluation and the associated mandates changed how teachers looked at data and used it to drive their instruction. Teacher leader Fred pointed to positive impact. He said that he looks at teaching and learning data significantly differently as a result of the connection to his measured effectiveness being tied to the data.

Implications of the Findings of This Study

Implications for practitioners. Findings and conclusions from this study could assist educators in identifying strategies and patterns that may be identified in major change initiatives, particularly those imposed by legislation beyond the local level. Of particular note, establishing a culture of trust among stakeholders in an organization appears to be critical to accomplishment of goals. Establishing joint ownership in regard to data appeared to be centralized in this study.

Darling-Hammond (2013) reported that anxiety related to determining who is and who is not an effective teacher is one of the most pressing topics in education.

Findings in this study implied that anxiety relative to teacher evaluation can be anticipated and ameliorated to some extent by fostering teacher–administrator relationships and dialogue about data-driven professional practice and student outcomes.

Tucker and Stronge (2005) advocated for the inclusion of student achievement data to be integrated into the evaluation process. Later studies of Toch and Rothman (2008) and Weisberg, Sexton, Mulhern, and Keeling (2009) found that the evaluative processes in the country as a whole were not designed to promote educational improvement.

Teachers interviewed in this study acknowledged that, as a result of the experience encountered in complying with state mandates to link teacher evaluation and student
achievement, they looked more closely at student data in search of the whole picture of student growth. They also credited at least a portion of the success the district has achieved to the integration and evolution of the Danielson rubric and the incorporation of student achievement data. These findings implied that practitioners could benefit from reflection about ways that student data could lead to improved instruction and student achievement.

Implications for administration. Pilot studies of student growth tools conducted in Michigan found that most districts lacked policy to address PA 102 (2011). Although principals reported the equivalent of 31 full work days conducting evaluations, only 60% of principals felt confident in this task or the vendor-provided tool, citing minimal training in respect to implementing the tool. Further, findings of primary importance from the piloting process showed that 89% of principals believed that the evaluation tools they tested focused on integral aspects of teaching, whereas only 62% of teachers believed that ratings assigned to them were accurate, and only 47% of teachers reported the tools easy to understand. These findings strongly suggest that professional development be included as a significant element in preparation of administrators to lead the implementation process in development of an integrated teacher evaluation program.

Administrators in this study demonstrated that establishing relationships of trust with teachers provided a firm foundation for coping with the mandated change in the teacher evaluation process. Administrative leadership was shown in time spent in dialogue with staff about expectations and guidance for understanding and applying new tools to integrate teacher evaluation and student achievement data. It was within the scope of these conversations where administrators felt that they were having an impact on student achievement, as the conversations accelerated opportunities to discuss pedagogical practice.
Having a research-proven instrument that is predicated on teacher improvement provided a blueprint for these discussions. Administrators’ collaborative approach and flexibility to adapt to change provided confidence needed to reduce anxiety among staff.

Approaching change at the top of an educational organization with a business mentality helped to bring stakeholders together in an intentional effort to comply with the law. This is particularly effective when a team of established educators embody the characteristics that Rogers and Shoemaker (1971) identified in early adopters.

**Implication for policy-makers.** According to Clotfelter (2006), accountability in respect to student achievement data is an important topic that continues to drive policymakers. Strom (2011) surveyed parents and found that it was an expectation that the state would ensure that high standards are in place and maintained for schools and teachers. Walters (2008) said that society expects the return on the investment in public education to be equitable in terms of allocated resources. He contended that districts are stewards of significant resources and the demand for maximizing the value and utility of these resources must be quantified.

Policy-makers have enacted federal and state legislation to initiate major change in approaches to teacher evaluation and student assessment. No Child Left Behind Act (NCLB, 2001), a federal initiative, aimed at increasing academic achievement as the focus of school improvement and required that all students and sub-groups of students be assessed annually. State legislation influenced the employment status of teachers in mandated requirements to link teacher evaluation and student achievement.

In Michigan, the challenge was to determine what system-wide student assessment best correlated with teacher evaluation methods. Led by the University of Michigan,
Institute for Social Research in 2013, the Governor’s Council for Educator Effectiveness initiated a pilot of four observation tools that included Danielson’s framework for teaching (1996), 5 Dimensions of Teaching and Learning (CEL, 2007), Marzano’s teacher evaluation model (2012), and Thoughtful Classroom framework (Silver Strong, 2007). Four student growth tools were also piloted in Michigan: (a) Northwest Evaluation Association (NWEA) for Grades K-6 (b), ACT Explore for Grades 7-8, (c) ACT Plan for Grades 9-10, and (d) ACT for Grade 12.

The Michigan school district leaders and teachers familiar with the evaluation alterations and protocols and the Department of Education [MDE] benefitted by meeting together. MDE contracted with a professional research company to gather information on experiences of teachers and administrators to assist in evaluating the effectiveness of educator evaluation systems in Michigan school districts. Their intent was to identify barriers to effective implementation and evaluate the effectiveness of measuring student growth using alternative methods. This findings of this study provided data about how one rural northern Michigan district responded and experienced the change process.

This case study provided school district policy-makers a greater depth of understanding about impact of change throughout the organization. Knowledge of the phenomenon that impacts the district when change is enacted and the potential implications associated with introducing or altering policy that relates to educator evaluation should inform the school board’s process of crafting policy and protocol.

Implications for the community. According to Walters (2008), legislation has dictated that educators must demonstrate to society a return on the investment into student learning. Yeh (2010) conducted a study of the cost-effectiveness of 22 various approaches to
education reform. He found that after eliminating variables, the rapid assessments of children for educational purposes garnered the best return on investment. Successful strategies employed by stakeholders in the school district to meet the challenge of mandated change were recognized by other school districts, state school leadership, union leaders, and policy-makers. Measures were instituted in the school district to achieve the integration of student growth data and individual teachers’ evaluations. Implications to the community include an appreciation of the collaborative efforts of the stakeholders in educational service in the community. Pride in accomplishments may translate into increased parental involvement and overall support for the school district.

**Personal implications for the researcher.** Findings of this study offered numerous implications for my role in educational leadership. As superintendent, I have developed a deeper understanding of four areas of focus. First, establishing trust by dedicating time for honest focused discussion on improved instructional practices produced through the evaluation process is critical. In the course of this research, I recognized how challenging a change initiative would be to navigate without the opportunity to work closely with individuals.

Second, I recognized that any change initiative will result in people adapting differently but within certain contextual groups. My perspective on the responsiveness of individuals to respond to change has grown to understand that people will respond at different paces. Review of Rogers and Shoemaker’s (1971) adopter types helped me to recognize that individuals will respond to change in a personal way.

Third, the inherent value associated with sense-making for individuals involved in a phenomenon is critical to understanding the impact of change. Without the opportunity to
analyze the changes associated with the process in this study, I would not have realized how change can be deeply personal. It was through the concept of sense-making that I recognized that the impact of sweeping reform impacts individuals differently even if they are in the same basic employment classification. These changes have the capacity to involve different characteristics and attributes in unique ways based on prior experiences.

Finally, understanding qualitative research procedures will assist me in introducing and evaluating educational programs in the future. A conscious objective of personalizing the implications to each colleague will be a more effective approach to integrating programs and processes in an educational setting. Providing forums for staff to reflect on professional practice and how it is evolving will help to guide the way to optimizing student performance.

I have adopted a new perspective and appreciation for the value of establishing relationships predicated on trust and the expenditure of a principal’s time. Sound qualitative research yielded greater insight into the impact of change. Through these experiences, I will be better poised to continue my growth as an educator and to assist my colleagues in a more effective position to serve students.

**Recommendations for Future Research**

This study served to provide educational stakeholders and policymakers with a depiction of how one northern Michigan rural school district has responded to state-mandated changes in teacher evaluation. The context of the district studied provided a setting where the size of the district seemed ideal for providing opportunities for growth. It would be challenging to understand how larger setting districts prepared, enacted, and sustained the strategies necessary for the integration of Public Act 102 (2011). In an urban or suburban setting, it could prove challenging to accommodate the time necessary for a principal and
other district administrators to focus as much attention on evaluation as the process as was shown in the present study.

The Michigan Department of Education continues to identify best practices, barriers to effectiveness, and most effective student growth measures. Comparing and contrasting the experiences of districts in different locations could contribute to a broader perspective about the integration of teacher evaluation and student achievement data.

Legislation stipulates that for 2018–2019, the basis of teacher evaluation will be 40% of student growth. Continuing research about progress toward the integration of teacher evaluation and student growth data may indicate whether efforts show the anticipated increase in student achievement and explore the efficacy of plans to tighten the linkage.

The State of Michigan has currently continued to allow individual districts to autonomously identify what data to consider in teacher evaluation. Research about whether this is an appropriate measure based on teachers’ perspective seems logical. Understanding whether districts are prepared and confident in the metrics that they are using to ascertain student growth could yield valuable insight into how confident teachers are toward accepting teacher evaluation as an accurate, valid, and reliable process.

Future studies include possibility of including student, colleague, and parental perspectives as potential sub-sections of an evaluation process. Although inclusion of these perspectives should refrain from being punitive due to reliability issues, there may be interest in expanding the number and categories of respondents who have a voice in the process of teacher evaluation.
Summary

The purpose of this study was to identify how one district attempted to meet the need for greater accountability in the teacher evaluation process as mandated by law and to describe strategies that were employed by the district to ensure the integration of student achievement data.

The study was developed on a conceptual framework that emphasized the process of change and the responses of the stakeholders to change within an organization. The concept of sense-making, as described by Choo (2006), explained the importance of understanding the personal orientation and growth achieved in the experience and interpretation of a phenomenon by those involved. A comprehensive review of relevant literature provided an historical foundation and added substance to the text describing key aspects of the investigation.

A qualitative, single case study design included interviews with 17 stakeholders representing various levels of participation and responsibility. Ten themes that evolved from analysis of the data gathered in the interviews became the major findings in this research. The themes represented a composite of the efforts of the school district to comply with mandated state legislation, revealed a consensus of the meaning interpreted by the personal experience of the stakeholders, and offered suggestions for future research.

Major change affects the culture of an organization and all stakeholders within. The stakeholders in one rural northern Michigan school district accepted the challenge of integrating student achievement data into a new process of teacher evaluation. The process of change continues along with social, economic, and political concerns, but one district in
Michigan remains committed to work together toward the worthy objective of improving student achievement.
References


Diriwächter, R., & Valsiner, J. (January 2006) Qualitative development research methods in their historical and epistemological contexts. *FQS. 7*(1), Art. 8.


Appendix
Appendix A: UHSRC Approval

RESEARCH @ EMU

UHSRC Determination:  EXPEDITED INITIAL APPROVAL

DATE: January 3, 2017

TO: Scott Moore, B.S., MPA, Ed.S.
    Eastern Michigan University

Re: UHSRC: # 955668-1
Category: Expedited category 7
Approval Date: January 3, 2017
Expiration Date: January 2, 2018


Your research project, entitled A Case Study of the Design and Implementation of a New Teacher Evaluation System in a Northern Michigan School District, has been approved in accordance with all applicable federal regulations.

This approval included the following:

1. Enrollment of 16 subjects to participate in the approved protocol.
2. Use of the following study measures:

   Teacher Leader Interview Protocol; Administrator and Former Superintendent Interview
   Protocol; Board Member Interview Protocol; Technology Director Interview Protocol
3. Use of the following stamped recruitment materials: Educator Letter
4. Use of the stamped: Informed Consent form

Renewals: This approval is valid for one year and expires on January 2, 2018. If you plan to continue your study beyond January 2, 2018, you must submit a Continuing Review Form by December 3, 2017 to ensure the approval does not lapse.

Modifications: All changes must be approved prior to implementation. If you plan to make any minor changes, you must submit a Minor Modification Form. For any changes that alter study design or any study instruments, you must submit a Human Subjects Approval Request Form. These forms are available through IRBNet on the UHSRC website.

Problems: All major deviations from the reviewed protocol, unanticipated problems, adverse events, subject complaints, or other problems that may increase the risk to human subjects or change the category of review must be reported to the UHSRC via an Event Report form, available through IRBNet on the UHSRC website

Follow-up: If your Expedited research project is not completed and closed after three years, the UHSRC office requires a new Human Subjects Approval Request Form prior to approving a continuation beyond three years.

Please use the UHSRC number listed above on any forms submitted that relate to this project, or on any correspondence with the UHSRC office.

Good luck in your research. If we can be of further assistance, please contact us at 734-487-3090 or via e-mail at human.subjects@emich.edu. Thank you for your cooperation.

- 1 -
Appendix B: Invitation to Participate in the Study

Dear Educator,

I am conducting research in an effort to complete a doctoral thesis in educational leadership. I am writing to ask you to consider participating an interview to be used for my dissertation as part of my doctoral program at Eastern Michigan University. I have chosen to conduct research about the teacher evaluation process in Michigan. The purpose of this research is to understand the impact of the changes in the law from 2010 to the present.

Participation in this research is voluntary and involves participating in an interview. The interview should take approximately 60 minutes to complete and I will be in contact with you as a follow-up to schedule the interview.

Your responses will be anonymous and only a code number will identify your response. At no time will your name be associated with your responses to the interview. All information will be secure at all times and kept completely confidential. Electronic data will be stored on a password protected computer and printed data will be stored in a locked file cabinet. Only the principal investigator and dissertation chair will have access to the complete set of data.

There are no known risks to you by participating in the survey. Once you have completed the interview, you will be asked to review your responses for authenticity. Upon your validation, your participation will have been completed. You may choose not to participate. You can change your mind at any time and withdraw from this research without negative consequences.

Results will be provided in aggregate form only. No names or individually identifying information will be revealed. There are no known benefits to you as a result of participating in this research other than the findings may be found to be valuable to the field of education. The results may be presented at conferences, in scientific publications, and as part of a doctoral thesis being conducted by the principal researcher.

You will be asked to sign an "Informed Consent Form" indicating your consent to participate in this research. If you have any questions about the research, you can contact the Principal Investigator, Scott Moore at smoore64@emich.edu, by phone at 989-820-6718 or the dissertation chair, Dr. Ronald Williamson at rwilliams1@emich.edu.

For questions about your rights as a research subject, you can contact the Eastern Michigan University Office of Research Compliance at human.subjects@emich.edu or by phone at 734-487-3090.

Thank you in advance for your participation in this research study.

Sincerely,

Scott Moore, Principal Investigator

Approved by the Eastern Michigan University Human Subjects Review Committee
UHRC Protocol Number: 955668-1
Study Approval Dates: 01/03/17 – 01/02/18
Appendix C: Teacher Leader Interview Protocol

Date: ________________________________
School: ________________________________
Name of Interview Subject(s): ________________________________
Start Time: _______ End Time: _______ Total Time: ________

[Introduction]

Thank you for taking your time to meet with me. I am working with Eastern Michigan University to complete my research study and dissertation on how one Northern Rural Michigan District has responded to changes in teacher evaluation. Therefore, the purpose of this interview is to learn more about your experiences with the teacher evaluation tool and process at your school. So that I do not miss any of your comments, I would like to tape record our discussion. I have asked for your permission to this, as it will make my research work much easier. I should point out that your contribution will be anonymous and confidential, and that any published research will contain changed names. This interview should take around 60 minutes. Do you have any questions for me before we begin?

1. Please tell me what your role with the district is and has been since the summer of 2010.
2. Please relate to me what you can recall in regards to the history of teacher evaluation at the District and to what degree has improvement been made?
3. Was the emphasis on student achievement data compatible with how the district and you previously analyzed student results prior to 2010?
4. In your opinion, how complex and challenging were the changes in evaluation process for you to understand and in-turn utilize? Do you feel that you had a role in crafting and explaining the process to your colleagues?
5. Do you think the evaluation tool has contributed to enhanced school improvement and higher student achievement to a greater, lesser or the same degree than what you felt it would have made an impact at its introduction in 2010.
6. What do you think about the use of additional data sources in the evaluation model, such as peer review, student/parent feedbacks? Is this something that should be explored on a limited basis and evaluated?
7. What do you see as the next steps for ensuring sustainability of a teacher evaluation model into the future?
8. Additional Comments

[Closing]

Thank you for taking your valuable time to answer these questions. Your comments and answers will be invaluable for our research study. I will keep you in contact throughout the study with verification of this interview transcription and also at the end of this study, you will receive a final copy of the report.
Appendix D: Administrator and Former Superintendent Interview Protocol

Date: _________________________________________
School: _________________________________
Name of Interview Subject(s):________________________________________
Start Time: _______ End Time: _______ Total Time: __________

[Introduction]

Thank you for taking your time to meet with me. I am working with Eastern Michigan University to complete my research study and dissertation on how one Northern Rural Michigan District has responded to changes in teacher evaluation. Therefore, the purpose of this interview is to learn more about your experiences with the teacher evaluation tool and process at your school. So that I do not miss any of your comments, I would like to tape record our discussion. I have asked for your permission to this, as it will make my research work much easier. I should point out that your contribution will be anonymous and confidential, and that any published research will contain changed names. This interview should take around 60 minutes. Do you have any questions for me before we begin?

1. Please tell me what your role with the district is and has been since the summer of 2010.
2. Please relate to me what you can recall in regards to the history of teacher evaluation at the District and to what degree has improvement been made?
3. Was the emphasis on student achievement data compatible with how the district and you previously analyzed student results prior to 2010?
4. In your opinion, how complex and challenging were the changes in evaluation process for you to understand and in-turn utilize? Do you feel that you had a role in crafting and explaining the process to your colleagues?
5. Do you think the evaluation tool has contributed to enhanced school improvement and higher student achievement to a greater, lesser or the same degree than what you felt it would have made an impact at its introduction in 2010.
6. What do you think about the use of additional data sources in the evaluation model, such as peer review, student/parent feedbacks? Is this something that should be explored on a limited basis and evaluated?
7. What do you see as the next steps for ensuring sustainability of a teacher evaluation model into the future?
8. Did you find that the Modified-Danielson Evaluation Rubric was more beneficial to driving teacher instructional practices or do you feel that the data procured from InFORMED was more beneficial to the instruction of your staff? Please explain.
9. If you could have improved any aspect of the teacher evaluation system, what would it be and why?
10. Additional Comments

[Closing]

Thank you for taking your valuable time to answer these questions. Your comments and answers will be invaluable for our research study. I will keep you in contact throughout the study with verification of this interview transcription and also at the end of this study, you will receive a final copy of the report.
Appendix E: Board Member Interview Protocol

Date: _________________________________________
School: _________________________________
Name of Interview Subject(s):________________________________________
Start Time: _______ End Time: _______ Total Time: __________

[Introduction]
Thank you for taking your time to meet with me. I am working with Eastern Michigan University to complete my research study and dissertation on how one Northern Rural Michigan District has responded to changes in teacher evaluation. Therefore, the purpose of this interview is to learn more about your experiences with the teacher evaluation tool and process at your school. So that I do not miss any of your comments, I would like to tape record our discussion. I have asked for your permission to this, as it will make my research work much easier. I should point out that your contribution will be anonymous and confidential, and that any published research will contain changed names. This interview should take around 60 minutes. Do you have any questions for me before we begin?

Board Members:
1. Please tell me what your role with the district is and has been since the summer of 2010.
2. What direction did the Board give for the implementation of Public Act 102?
3. Please relate to me what you can recall in regards to the history of teacher evaluation at the District and to what degree has improvement been made?
4. Was the emphasis on student achievement data compatible with how the district and you previously analyzed student results prior to 2010?
5. In your opinion, how complex and challenging were the changes in evaluation process for you to understand and in-turn utilize? Do you feel that you had a role in crafting and explaining the process to your colleagues?
6. Do you think the evaluation tool has contributed to enhanced school improvement and higher student achievement to a greater, lesser or the same degree than what you felt it would have made an impact at its introduction in 2010.
7. What do you think about the use of additional data sources in the evaluation model, such as peer review, student/parent feedbacks? Is this something that should be explored on a limited basis and evaluated?
8. What do you see as the next steps for ensuring sustainability of a teacher evaluation model into the future?
9. What role have you and the Board of Education had in regards to constructing, implementing and altering the teacher evaluation tool employed by the district? Why do you believe that this level of involvement has taken place?
10. Additional Comments

[Closing]
Thank you for taking your valuable time to answer these questions. Your comments and answers will be invaluable for our research study. I will keep you in contact throughout the study with verification of this interview transcription and also at the end of this study, you will receive a final copy of the report.
Appendix F: Technology Director Interview Protocol

Date: _______________________________
School: _______________________________
Name of Interview Subject(s): _______________________________
Start Time: _______ End Time: _______ Total Time: __________

[Introduction]

Thank you for taking your time to meet with me. I am working with Eastern Michigan University to complete my research study and dissertation on how one Northern Rural Michigan District has responded to changes in teacher evaluation. Therefore, the purpose of this interview is to learn more about your experiences with the teacher evaluation tool and process at your school. So that I do not miss any of your comments, I would like to tape record our discussion. I have asked for your permission to this, as it will make my research work much easier. I should point out that your contribution will be anonymous and confidential, and that any published research will contain changed names. This interview should take around 60 minutes. Do you have any questions for me before we begin?

1. Please tell me what your role with the district is and has been since the summer of 2010.
2. What direction did the Board give for the implementation of Public Act 102?
3. Please relate to me what you can recall in regards to the history of teacher evaluation at the District and to what degree has improvement been made?
4. Was the emphasis on student achievement data compatible with how the district and you previously analyzed student results prior to 2010?
5. In your opinion, how complex and challenging were the changes in evaluation process for you to understand and in-turn utilize? Do you feel that you had a role in crafting and explaining the process to your colleagues?
6. Do you think the evaluation tool has contributed to enhanced school improvement and higher student achievement to a greater, lesser or the same degree than what you felt it would have made an impact at its introduction in 2010.
7. What do you think about the use of additional data sources in the evaluation model, such as peer review, student/parent feedbacks? Is this something that should be explored on a limited basis and evaluated?
8. What do you see as the next steps for ensuring sustainability of a teacher evaluation model into the future?
9. How much time annually have you had to dedicate to collecting and chronicling student-achievement data for InFORMED and has this changed over-time?