Investigating Practices of Research-Proven Multidimensional Teacher Evaluation Systems in Michigan Schools

George A. Aramath
Western Michigan University, garamath@gmail.com

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Leadership Commons, and the Teacher Education and Professional Development Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/233

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
INVESTIGATING PRACTICES OF RESEARCH-PROVEN MULTIDIMENSIONAL TEACHER EVALUATION SYSTEMS IN MICHIGAN SCHOOLS

by

George A. Aramath

A dissertation submitted to the Graduate College in partial fulfillment of the requirements for the degree of Doctor of Philosophy Educational Leadership, Research and Technology Western Michigan University April 2014

Doctoral Committee:
Joseph Kretovics, Ph.D., Chair
Patricia Reeves, Ed.D.
Mattie Hampton, Ph.D.
INVESTIGATING PRACTICES OF RESEARCH-PROVEN MULTIDIMENSIONAL TEACHER EVALUATION SYSTEMS IN MICHIGAN SCHOOLS

George A. Aramath, Ph.D.

Western Michigan University, 2014

Traditional teacher evaluation systems have little to no impact on changing teacher behavior or improving student achievement. Subsequently, studies and reports show the lack of and need for multidimensional teacher evaluation system. This deficiency and need is especially evident in Michigan due to the state's recent legislative mandate that their evaluation system must include multiple data as measures of educator effectiveness.

The purpose of this study is to investigate and describe how two Michigan public schools are attempting to build a multidimensional teacher evaluation system that includes the new Michigan mandate to incorporate evidence of student learning as a significant element in teacher performance review systems.

This study makes use of a qualitative case study design that includes the following sources: interviews of school personnel, review of archival documents and teacher surveys, and observations of the implementation process. The investigation answers five research questions: What strategies and processes are used by these schools in Michigan to build research-supported multidimensional teacher evaluation systems based on Charlotte Danielson's Framework for Teaching? How are schools addressing each of Danielson's four domains of teacher practice? What resources are utilized to
implement a multidimensional teacher evaluation system? What challenges do schools face in implementing a multidimensional teacher evaluation system and how are they addressed? What evidences exist that the teacher evaluation system is resulting in positive educational outcomes?

National and state reports acknowledge that there are few comprehensive models that exist in meeting Michigan's mandate to build a rigorous, transparent, and fair performance evaluation system. This research study addresses the need to identify and describe promising practices so that it will assist state policy makers, local districts, and individual schools to assess their current evaluation practices in order to determine possible strategies for improvement.

An investment to provide all students with a quality education, which includes ensuring that they are taught by a highly effective teachers, has the potential to positively transform the lives of students and communities.
ACKNOWLEDGMENTS

This journey to a foreign terrain would not happen without my guides. It began with the support and direction of my dissertation committee. As my primary mentor and guide, Dr. Joseph Kretovics encouraged me in this path towards completing my doctorate. His ideas and thoughts sustained my dissertation writing. And for this, I am grateful to him. I thank Dr. Patricia Reeves for serving on my committee. Her suggestions helped narrow my subject and methodology. Plus, her encouragement and enthusiasm for the topic propelled me further into my research study. I also thank Dr. Mattie Hampton for serving on my committee and offering practical suggestions. Besides my committee members, I thank Dr. Poppink and Dr. Bierlein Palmer for their perspectives that made my dissertation richer. Lastly, I thank the two schools who opened their doors to this study.

My parents, Alias and Elizabeth Aramath and my second parents, Raj and Annamma Perumala have always been supportive of my educational aspirations. Thank you for your encouragement and prayers.

I want to especially thank my wife, Sherley. From the very first step, she has been at my side when I applied for the doctoral program and when I submitted my final copy of the dissertation. Without her, I could not have completed this journey. Now I look forward to taking other journeys with my wife and children! And most importantly, I thank God Almighty who makes all things possible.

George A. Aramath
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .............................................................................................................................. ii

LIST OF TABLES ..................................................................................................................................... ix

LIST OF FIGURES .................................................................................................................................... xi

CHAPTER

I. INTRODUCTION ...................................................................................................................................... 1

  Background ............................................................................................................................................. 2

  Problem Statement ............................................................................................................................... 6

    Practical Problem Statement .............................................................................................................. 6

    Research Literature Problem Statement .......................................................................................... 8

  Purpose Statement ............................................................................................................................... 10

  Research Questions ............................................................................................................................ 11

  Literature Review ............................................................................................................................... 12

  Significance of Study ............................................................................................................................ 12

  Methods Overview ............................................................................................................................. 14

  Summary .............................................................................................................................................. 15

II. LITERATURE REVIEW ...................................................................................................................... 17

  Key Background Information ............................................................................................................ 17

    NCLB and Highly Qualified Teachers .............................................................................................. 17

    Impact of Teacher on Student Achievement .................................................................................. 20
<table>
<thead>
<tr>
<th>Chapter Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Evaluation System</td>
<td>23</td>
</tr>
<tr>
<td>Purpose and Objective of Teacher Evaluation</td>
<td>24</td>
</tr>
<tr>
<td>Deficiencies in Traditional Teacher Evaluation Systems</td>
<td>28</td>
</tr>
<tr>
<td>Focus of the Study</td>
<td>32</td>
</tr>
<tr>
<td>Multidimensional Teacher Evaluation System</td>
<td>32</td>
</tr>
<tr>
<td>Federal and State of Michigan Policy</td>
<td>36</td>
</tr>
<tr>
<td>Effective Teaching</td>
<td>39</td>
</tr>
<tr>
<td>Definition of teacher quality and effective teaching</td>
<td>39</td>
</tr>
<tr>
<td>Danielson’s framework for teaching</td>
<td>42</td>
</tr>
<tr>
<td>Components of Multidimensional Teacher Evaluation System</td>
<td>52</td>
</tr>
<tr>
<td>Principal’s Role</td>
<td>56</td>
</tr>
<tr>
<td>Student Achievement Data</td>
<td>63</td>
</tr>
<tr>
<td>Teacher Portfolio</td>
<td>72</td>
</tr>
<tr>
<td>Student Evaluation of Teachers</td>
<td>78</td>
</tr>
<tr>
<td>Peer Review</td>
<td>84</td>
</tr>
<tr>
<td>Other Instruments</td>
<td>92</td>
</tr>
<tr>
<td>Teacher self-reflection</td>
<td>92</td>
</tr>
<tr>
<td>Analysis of lesson plans and classroom artifacts</td>
<td>94</td>
</tr>
<tr>
<td>Parent feedback</td>
<td>96</td>
</tr>
<tr>
<td>Conclusion</td>
<td>98</td>
</tr>
</tbody>
</table>
# Table of Contents—Continued

## CHAPTER

**III. METHODOLOGY** .................................................................................................................. 104
- Research Design ......................................................................................................................... 104
- Sampling, Subjects and Setting .................................................................................................. 107
- Data Collection Methods, Procedures and Instrumentation ..................................................... 110
  - Document Analysis .................................................................................................................... 111
  - Individual Interview .................................................................................................................. 113
  - Focus Group .............................................................................................................................. 116
  - Observation ............................................................................................................................... 118
  - Online Survey .......................................................................................................................... 120
- Data Collection Phases .............................................................................................................. 121
- Data Analysis Processes and Procedures .................................................................................. 123
- The Researcher .......................................................................................................................... 124
- Delimitations .............................................................................................................................. 126

**IV. FINDINGS** .......................................................................................................................... 128
- Overview of Findings .................................................................................................................. 129
  - Data Analysis Processes and Procedures ................................................................................. 130
- Overview of Survey Results ....................................................................................................... 134
- Research Findings: Two Case Schools ...................................................................................... 136
  - Case School #1 ......................................................................................................................... 136
  - Background of the School ........................................................................................................ 136
  - Goal of the Teacher Evaluation System .................................................................................. 137
**Table of Contents—Continued**

**CHAPTER**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation of the Teacher Evaluation System</td>
<td>138</td>
</tr>
<tr>
<td>History</td>
<td>138</td>
</tr>
<tr>
<td>Description of teacher evaluation system</td>
<td>140</td>
</tr>
<tr>
<td>Danielson’s domains and evaluation instruments</td>
<td>147</td>
</tr>
<tr>
<td>Survey Results: Case School #1</td>
<td>148</td>
</tr>
<tr>
<td>Research Questions Examined</td>
<td>154</td>
</tr>
<tr>
<td>Question #1: Danielson framework</td>
<td>154</td>
</tr>
<tr>
<td>Question #2: Resources</td>
<td>156</td>
</tr>
<tr>
<td>Question #3: Challenges</td>
<td>158</td>
</tr>
<tr>
<td>Question #4: Evidence of impact</td>
<td>163</td>
</tr>
<tr>
<td>Case School #2</td>
<td>165</td>
</tr>
<tr>
<td>Background of the School</td>
<td>165</td>
</tr>
<tr>
<td>Goal of the Teacher Evaluation System</td>
<td>166</td>
</tr>
<tr>
<td>Implementation of the Teacher Evaluation System</td>
<td>167</td>
</tr>
<tr>
<td>History</td>
<td>167</td>
</tr>
<tr>
<td>Description of teacher evaluation system</td>
<td>168</td>
</tr>
<tr>
<td>Survey Results: Case School #2</td>
<td>173</td>
</tr>
<tr>
<td>Research Questions Examined</td>
<td>178</td>
</tr>
<tr>
<td>Question #1: Danielson framework</td>
<td>178</td>
</tr>
<tr>
<td>Question #2: Resources</td>
<td>182</td>
</tr>
<tr>
<td>Question #3: Challenges</td>
<td>184</td>
</tr>
</tbody>
</table>
Table of Contents—Continued

CHAPTER

Question #4: Evidence of impact.............................................. 191

V. DISCUSSION AND CONCLUSION ....................................................... 195

Discussion of Results........................................................................ 195

Research Question #1: Danielson Framework................................. 196
Research Question #2: Resources.................................................... 198
Research Question #3: Challenges.................................................. 200
Research Question #4: Evidence of Impact .................................... 201

Recommendations............................................................................ 202

Planning and Preparation ................................................................. 203
Branding and Marketing of the New Evaluation Model..................... 204
Professional Development ................................................................. 206
Principal’s Role................................................................................ 208
Differentiation .................................................................................. 210

Ideas for Future Research ............................................................... 212

REFERENCES ..................................................................................... 216

APPENDICES

A. Interview Protocol......................................................................... 250
B. Focus Group Protocol .................................................................. 253
C. Evaluation Implementation Observation Protocol........................ 256
D. Professional Development Observation Protocol......................... 258
E. Portfolio Artifact Matrix ............................................................... 260
Table of Contents—Continued

APPENDICES

F. Online Survey .................................................................................................................. 263

G. Coaching Observation Form .......................................................................................... 266

H. Professional Reflection Sheet ....................................................................................... 269

I. Video Reflection Form ................................................................................................... 271

J. Lesson Study Template ................................................................................................. 273

K. Lesson Observation Report .......................................................................................... 275

L. Teacher Self-Checklist ................................................................................................... 278

M. Summative Evaluation Form: Indicators of Professional Practice ....................... 280

N. Danielson Rubric .......................................................................................................... 284

O. Individual Development Plan ...................................................................................... 307

P. HSIRB Approval Letter ................................................................................................. 310
LIST OF TABLES

1. Framework for Teaching................................................................. 45
2. Danielson Framework Matched With Evaluation Instruments.............. 53
3. Data Sources and Research Questions.............................................. 130
4. Data Analysis Plan............................................................................. 131
5. Demographic Profile of High School (2011-2012).............................. 137
6. Danielson’s Domains and Evaluation Instruments............................ 147
8. Distribution of New and Veteran Teachers: Case School #1 ............... 149
9. Statements Categorized Under School Addressing Danielson’s Domains (Research Question #1).............................................................. 150
10. Statements Categorized Under Use of Resources (Research Question #2) ...... 151
11. Statements Categorized Under Challenges Encountered (Research Question #3)................................................................. 151
12. Statements Categorized Under Evidence of Growth (Research Question #4) .... 152
13. Demographic Profile of Middle School (2011-2012).......................... 166
14. Middle School Evaluation System: Data Component ......................... 170
15. Distribution of New and Veteran Teachers: Case School #2 ............... 173
16. Statements Categorized Under School Addressing Danielson’s Domains (Research Question #1).............................................................. 174
17. Statements Categorized Under Use of Resources (Research Question #2) ...... 175
List of Tables—Continued

18. Statements Categorized Under Challenges Encountered (Research Question #3) ........................................................................................................................................ 176

19. Statements Categorized Under Evidence of Growth (Research Question #4) .. 177
LIST OF FIGURES

1. Components of teacher quality ................................................................. 40
2. Coding process for inductive analysis ....................................................... 133
3. Flowchart of two tracks of high school evaluation system ....................... 142
4. Flowchart of two tracks of middle school evaluation system .................... 172
5. Sample multidimensional evaluation model .............................................. 212
CHAPTER I
INTRODUCTION

In the landmark Supreme Court ruling that removed segregation and inequality in public education, Chief Justice Warren wrote: “Education is perhaps the most important function of state and local governments. . . In these days, it is doubtful that any child may reasonably be expected to succeed in life if he is denied the opportunity of an education” (Brown v. Board of Education, 1954, para. 16). Echoing similar sentiments, John Dewey was quoted as saying: “‘Education is not preparation for life; education is life itself’” (as cited in Littky & Grabelle, 2004, p. xviii). With this understanding of the significant influence of education, increased focus at the federal, state, and local level in public education is currently placed on improving student achievement so that all students will have the opportunity to a fair and equal education. Teachers play an important role in achieving this goal. As Darling-Hammond (1997a) stated, education is grounded on one of the fundamental presumption that teaching matters. Beerens (2000) further stated, “Development of the teacher as a professional is key to a successful learning culture within a school” (p. xii). In fact, Wise, Darling-Hammond, McLaughlin, and Bernstein (1984) pointed out that researchers acknowledge that school districts have a responsibility to “hold teachers accountable to standards of practice that compel them to make appropriate instructional decisions on behalf of their students” (p. 80). Research has consequently given justifiable arguments that a key means towards significantly improving student outcomes is raising the level of teacher quality since teaching has a greater impact on student success than any other school-based factor (Wright, Horn, & Sanders, 1997). But the question that follows is how. A number of strategies have been
undertaken by reformers to improve public school teaching, such as licensing exams, performance pay, and alternative certification. Another valuable and powerful tool is teacher evaluations. Though teacher evaluations hold many purposes such as ensuring accountability and determining employment status, its ultimate goal is to improve teaching and learning. In fact, Toch and Rothman (2008) pointed out that teacher evaluations can be the means for school and teacher improvement, which is the core of the education enterprise. This qualitative case study will therefore examine research-supported teacher evaluation systems and frameworks that incorporate evidence of student learning as a significant element in teacher performance review systems.

**Background**

Various studies have shown that teachers have a strong influence on student achievement (Borman & Kimball, 2005; Hanushek et al., 2005; Nye, Konstantopoulos, & Hedges, 2004; Muijs, & Reynolds, 2002; Goldhaber, D. D., Brewer, D. J., & Anderson, D., 1999). With this understanding, the No Child Left Behind Act of 2001 (NCLB) spotlighted the connection between student achievement and teacher accountability (United States Department of Education, 2002). Student achievement became an important focus of this policy by mandating multiple assessments for students and monitoring performance of specific subgroups through annual measurable objectives known as adequate yearly progress. However, an equally important aspect of this legislation is the role of teachers who are now held accountable for their share in meeting student achievement goals. In particular, as a follow-up to NCLB, the American Recovery and Reinvestment Act (ARRA) of 2009 targeted a significant amount of funding towards improving state and local education systems. Commonly known as the
Race to the Top Fund, it provided $4.35 billion in competitive grants for states; these grants are based upon meeting certain criteria, one of which specifies that states should “design and implement rigorous, transparent, and fair evaluation systems for teachers . . . that differentiate effectiveness using multiple rating categories” (United States Department of Education, 2009, p. 9). For the state of Michigan, Public Act 205 was passed in 2009, known as “Race to the Top” School Reform, which required districts to conduct annual teacher evaluations that include student growth as a significant factor (S. Res. 981). In essence, these legislations at the federal and state level are meant to improve student achievement by, in part, focusing on improving teacher performance.

In regards to the connection between teachers and student learning, researchers have shown that instructional expertise of teachers is at the heart of the learning enterprise (Darling-Hammond, 2000; Wang, Haertel,, & Walberg, 1993). Consequently, Toch and Rothman (2008) wrote that teacher evaluation should be the center of the accountability portion of NCLB in order to measure teacher performance. Teacher evaluation should be an important focus because “without high quality evaluation systems, we cannot know if we have high quality teachers” (Stronge & Tucker, 2003, p. 3). After all, the center of education is the link between teaching and learning, and this relationship is strongest when effective teachers are educating students. While effectiveness will be later defined, in the end, researchers point to teacher evaluation as a vital mechanism for improving teaching and learning (Beerens, 2000; Borman & Kimball, 2005; Danielson & McGreal, 2000; Darling-Hammond, 2006a; Peterson, 2000). Danielson and McGreal (2000) quote from the writing of Donald Haefele to list the following purposes of a teacher evaluation system:
• Provide constructive feedback to individual educators
• Screen out unqualified persons
• Recognize and help reinforce outstanding service
• Provide direction for staff development practices
• Provide evidence that will withstand professional and judicial scrutiny
• Aid institutions in terminating incompetent or unproductive personnel
• Unify teachers and administrators in their collective efforts to educate children (p. 8).

In essence, quality teaching matters because as research has already shown, teachers influence what students learn. Therefore, by investing in improving teacher evaluation, student achievement is ultimately impacted.

Before looking at the current state of teacher evaluation systems, it is first important to define the term. Stufflebeam (2001) defined evaluation as “a study designed and conducted to assist some audience to assess an object’s merit and worth” (p. 11). More specifically, the Joint Committee on Standards for Educational Evaluation (2009) defines personnel evaluation as “the systematic assessment of a person’s performance and/or qualifications in relation to a professional role or some specified and defensible institutional purpose” (p. 3). Therefore, in this paper the term teacher evaluation will refer to the systematic assessment of teacher performance. A multidimensional teacher evaluation system incorporates multiple data sources and evaluators in the evaluation process. On the other hand, the term traditional teacher evaluation will refer to older evaluation systems that use a single or limited data source, typically completed by the school principal.
When examining teacher evaluation systems, one should first establish essential standards that encompass sound evaluation practice. Approved by the American National Standard Institute, the Joint Committee on Standards for Educational Evaluation (2009) established four essential attributes that are necessary and sufficient for a fair and sound evaluation system; these attributes include:

1. Propriety standards to ensure legal, ethical, and due regard for the welfare of the evaluatee and those involved in the evaluation.
2. Utility standards to guide evaluations so that they will be informative, timely, and influential.
3. Feasibility standards to ensure that the system is as easy to implement as possible, efficient in their use of time and resources, adequately funded, and viable from a political standpoint.
4. Accuracy standards to determine whether the evaluation has produced sound information and are technically adequate and as complete as possible (p. 6-7).

The unfortunate reality is that traditional teacher evaluation systems fail to fulfill both the above mentioned purposes and attributes of sound evaluation practices. 

Danielson and McGreal (2000) listed two of the main shortcomings that characterize traditional evaluation processes: “(1) outdated, limited evaluation criteria and (2) the lack of shared values and assumptions about what constitutes good teaching” (p. 11). Due to these shortcomings, traditional teacher evaluation systems turn out to be ineffective in keeping teachers accountable and improving their practice. Frase and Streshly (1994) summarized the state of teacher evaluation when they wrote, “Research and learned opinion strongly support the contention that teacher evaluation has been of little value”
Moreover, Loup, Garland, Ellett, and Rugutt’s (1996) study found that teacher evaluation practices in the 100 largest school districts in the United States showed little change during the past 10 to 15 years despite increased attention in educational accountability and school reform. In fact, a recent review by Learning Point Associates (2010) of the 41 states that applied for Race to the Top funding discovered that though these states made some changes in their evaluation system, half of the states still “do not have a student growth model, nor did they indicate current work leading to the development of a student growth model,” (p. 2) and only nine states reported that they currently differentiate teacher effectiveness using multiple rating categories. What’s more, one of the largest teacher’s union even admitted to the deficient state of teacher evaluation; the American Federation of Teachers (2011) began its recent press report by referring to the current state of teacher evaluation systems in the following manner: “With rare exceptions, teacher evaluation procedures are broken—cursory, perfunctory, superficial and inconsistent” (p. 1).

**Problem Statement**

**Practical Problem Statement**

The Joint Committee on Standards for Educational Evaluation (2009) listed the following major assumption behind evaluation systems: “The fundamental purpose of personnel evaluation must be to help provide effective services to students . . . personnel evaluation can and must be designed and constructed to encourage and guide evaluatees to perform more effectively” (p. 3). Unfortunately, traditional teacher evaluation systems fail to both adequately assess the complex and comprehensive scope of the teacher’s position and provide impact on professional growth. A recent example of these failures
comes from a study by the New Teacher Project of 12 districts in four states (Weisberg, Sexton, Mulhern, & Keeling, 2009). Their research revealed that more than 99% of teachers in districts using binary ratings were rated satisfactory whereas 94% obtained one of the top two ratings in districts using a broader range of ratings. Furthermore, “73 percent of teachers surveyed said their most recent evaluation did not identify any development areas” (Weisberg, Sexton, Mulhern, & Keeling, 2009, p. 9). Clearly, traditional evaluation systems are inadequate in assessing and developing teachers.

Additionally, states are only beginning to transition from a traditional to a more multidimensional teacher evaluation system. As an example, a 2007 report by the Institute of Education Sciences under the U.S. Department of Education examined the Midwest region’s district policies on teacher evaluation across 140 participating districts. The report found the following common theme: “Practices pertaining to the standards and criterion that inform the evaluations are less frequently referenced. Fewer than two in five districts provided details in their evaluation policy about the actual criteria used to rate teaching practice” (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007, p. 14). The specific criteria used within these districts also varied across these areas of teaching: knowledge and instruction, monitoring, professional responsibilities, classroom management, planning and preparation, and other. Finally, the same report discovered that most evaluation systems are not multidimensional, with school principals and administrators being the sole conductor and evaluator (Brandt et. al, 2007).

These types of findings laid the root for targeting a portion of the federal American Recovery and Reinvestment Act of 2009 to fund improvements in state and local education systems. Michigan specifically took primary steps to change their teacher
evaluation system by passing Public Act 205 in late 2009 which “evaluates a teacher’s or school administrator’s job performance, using multiple rating categories that take into account data on student growth as a significant factor” (S. Res. 981). Even though these types of legislative acts focused on improving teacher evaluations were recently passed in many states, especially with the prospect of obtaining Race to the Top federal funding, the National Council of Teacher Quality 2010 report that addressed key policy areas such as teacher preparation and evaluation “found that states had much work to do to ensure that every child has an effective teacher” (p. 3). In particular, their report for Michigan gave a grade of “D-” in the category of identifying effective teachers and an overall Michigan teacher policy grade of “D-” (NCTQ, 2010, p. 5). Though the report acknowledged Michigan’s recently passed legislations in this area, its section on critical attention area focused on “Michigan policies that need to better connect to teacher effectiveness” (NCTQ, 2010, p. 7). In summary, it has been established that traditional evaluation systems are inadequate in assessing and developing teachers, and states such as Michigan are only beginning to transition from a traditional to a more multidimensional teacher evaluation system. Therefore, this study will provide an alternative to traditional evaluation systems by examining the makeup of a multidimensional teacher evaluation system.

**Research Literature Problem Statement**

Though there is ample research showing the necessity for multidimensional teacher evaluation systems, there is a significant lack of research literature related to the process and implementation of this system. The Brookings Institute’s recent report on teacher evaluation expressed this deficiency by stating that there is an “immaturity of the
knowledge base on the design of teacher evaluation systems” (Glazerman et al., 2011, p. 1). Another study by the National Comprehensive Center for Teacher Quality further noted “the dearth of available research-based methods and models of comprehensive teacher evaluation” (Goe, Holdhelde, & Miller, 2011, p. 5). Due to the lack in methods and models of research-supported evaluation systems, the Education Alliance of Michigan (2011) in partnership with Michigan State University recently published a grant-funded analysis of Michigan teacher evaluation practices with literature on transformative multidimensional evaluation practices. This report provided recommendations on how school districts could begin moving toward establishing more growth oriented and dynamic evaluation systems with standards or research-based evaluation criterion that align with the state’s mandates on student achievement being a significant element in the evaluation of teachers. The report was prepared because “many districts and teachers’ unions asked for more guidance in designing plans that would be effective and appropriate for their districts” (Education Alliance of Michigan, 2011, p. 7). While it found that many school districts have yet to even adopt standards or research based evaluation instruments, the report mentioned that those that have are in a better position to move toward a more growth oriented or transformative model of teacher evaluation. In the end, the same report acknowledged that there are few comprehensive evaluation models that exist; so the recommendation was made for further studies to identify and describe promising practices and models. In addition, Michigan Association of School Personnel Administrators’ (2010) report similarly acknowledged a lack of guidance and literature on how to implement Michigan’s recent mandate for growth-oriented and dynamic evaluation systems. Hence, a gap exists in the literature related to
methods and models of research-supported evaluation models. Therefore, this study will attempt to partially fill this void by examining strategies and processes used by Michigan schools to build research-supported multidimensional teacher evaluation systems.

Purpose Statement

When studying the subject of teacher evaluation, it is first important to have a common definition and understanding of the complex activity of teaching. Danielson’s *Framework for Teaching* (2007) was originally developed in 1996 to describe various components of teacher’s daily work and underlying assumptions of the responsibilities of teaching. Irrespective of subject area or grade-level, this framework addresses the various characteristics of teaching through the use of four domains: planning and preparation, the classroom environment, instruction, and professional responsibilities. One of its key purposes is to serve as a guide for shared understanding and a common language of effective teaching. In other words, the framework helps to answer the question of what effective teachers know and do.

Studies and reports have already shown the lack of and the need for multidimensional teacher evaluation systems (Brandt et. al, 2007; Learning Point Associates, 2010; Weisberg, Sexton, Mulhern, & Keeling, 2009). This deficiency and need is especially evident in Michigan due to the state’s recent legislative mandate that their evaluation system must include multiple data as measures of educator effectiveness (Education Alliance of Michigan, 2011; Michigan Association of School Personnel Administrators, 2010; NCTQ, 2010; S. Res. 1509). The purpose of this research paper is therefore to describe how a criterion sample of Michigan schools that utilize Charlotte Danielson’s research-supported *Framework for Teaching* are attempting to build a
multidimensional teacher evaluation system that includes the new Michigan mandate to incorporate evidence of student learning as a significant element in teacher performance review systems.

It must be mentioned that though innovations in multidimensional teacher evaluation systems are beginning in Michigan schools, often the knowledge of these processes and practices is limited to the school site. This study will subsequently also help in disseminating information on how Michigan schools are moving toward developing multidimensional evaluation practices and systems.

**Research Questions**

The overarching research question that will guide this study is the following: What strategies and processes are used by schools in Michigan that are attempting to build research-supported multidimensional teacher evaluation systems based on Charlotte Danielson’s *Frameworks for Teaching*? From this overarching question, there will be four primary sub-questions:

1. How are schools addressing each of Danielson’s four domains of teacher practice as they begin to make their evaluation practices and systems more multidimensional and transformational (growth-oriented)?

2. What resources are utilized to implement a multidimensional teacher evaluation system?

3. What challenges do schools face in implementing a multidimensional teacher evaluation system and how are they addressed?

4. What evidence(s) exists that the teacher evaluation system has or is resulting in positive educational outcomes?
Literature Review

In describing the literature review, Creswell (2009) wrote, “Researchers use scholarly literature in a study to present results of similar studies, to relate that present study to the ongoing dialogue in the literature, and to provide a framework for comparing results of a study with other studies” (p. 45). This study therefore includes a review of literature related to the subject of teacher evaluation programs through reading similar studies on this subject, reflecting on what those studies found, and situating the study within the dialogue while also helping to substantiate the research problem.

Rudestam and Newton (2001) also wrote that the literature review will help connect the proposed study to prior work. In order to make this connection, key words were identified for searching the literature, which primarily took place through online databases such as ERIC, ProQuest, Google Scholar, and other more specialized databases and books with information related to teacher evaluation programs. While reading and collecting appropriate literature on the subject matter, I also considered and thereafter decided how to finally organize these studies into an overall structure.

Significance of the Study

Michigan schools are currently attempting to meet state standards on building a “rigorous, transparent, and fair performance evaluation system” (S. Res. 981, 2009). As the Education Alliance of Michigan’s (2011) and the Michigan Association of School Personnel Administrators’ (2010) report acknowledged, there are few comprehensive models that exist in meeting these requirements, and, therefore, further study is needed to identify and describe promising practices. This study will thus aid in partially filling this void in literature while also serving to assist state policy makers, local districts, and
individual schools in assessing their current evaluation practices to identify possible strategies for improvement.

As mentioned earlier, studies have shown that teachers have a strong influence on student achievement (Borman & Kimball, 2005; Nye, Konstantopoulos, & Hedges, 2004), while Toch and Rothman (2008) further pointed out that teacher evaluation should be the center of the accountability portion of quality teaching. Consequently, research studies focused on developing and improving teacher evaluation should ultimately impact student achievement such that all children will be afforded a quality education.

The value and significance of education should not be overlooked. Horace Mann (1872) reminded us, “If ever there can be a cause worthy to be upheld by all toil or sacrifice that the human heart can endure, it is the cause of education” (p. 7). Its impact is evident both at the individual and societal level. On an individual level, a quality education is the foundation for building productive citizens (Goldberg, 1983). On a societal level, a recent report by McKinsey and Company (2009) estimated that closing the achievement gap between White students and Latino and Black students increases United States’ annual Gross Domestic Product by more than half a trillion dollars. More specifically, Levin, Belfield, Muennig, and Rouse (2007) showed annual economic costs for the United States due to inequality in quality education for minority and low-income students; the study’s figures included state health loss of $11.6 billion, state crime related loss of $7.6 billion, and state tax losses of $40 billion. Clearly, education impacts different realms of society, the least of which includes employment, healthcare, incarceration, and civic participation. Accordingly, an investment to provide all students
with a quality education, which includes ensuring that they are taught by a highly effective teacher, can transform communities.

**Methods Overview**

This dissertation involves systematic steps and approaches in examining specific research questions. One of these approaches is known as qualitative research. Gall, Gall, and Borg (2007) wrote, “‘Qualitative researchers study things in their natural setting, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them’” (Gall, Gall, & Borg, 2007, p. 31). Since this study mainly takes place within the natural setting of schools in Michigan and involves reporting multiple perspectives of teacher evaluation models, a qualitative research approach through case studies is appropriate. Stake (1978) argued that the strength of case studies is its value in framing everyday experience through which one learns about the case firsthand. In other words, case studies provide a “full and thorough knowledge of the particular” (Stake, 1978, p. 7). Creswell (2009) further noted, “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. 13). In fact, the case study strategy of inquiry matches the overarching research question of this study, which examines processes used by schools in Michigan that are attempting to build research supported multidimensional teacher evaluation systems. Additionally, the subsequent research sub-questions go deeper into the examination of the evaluation process in school settings by focusing on resources, evidence, and challenges of schools in Michigan that develop a multidimensional model. As Marshall and Rossman (2011) wrote, “Case studies take the reader into the setting with vividness and detail not typically present in more analytic reporting formats” (p.
Yin (2003b) also explained that case studies are used “when the focus is on a contemporary phenomenon within some real-life context” (p. 1). Hence, the methodology of this study is a qualitative approach through a case study design that will ultimately address both the study’s purpose and research questions.

Summary

The United States of America has long defended the right to equal opportunity for all individuals. The Supreme Court in Brown v. Board of Education specifically acknowledged this right by connecting it to public education. An influential movement that began after this court ruling is the progression from a civil rights emphasis in educational policy that gave way “to an emphasis on accountability for educational outcomes” (Mitchell, Crowson, & Shipps, 2011, p. 5). The Supreme Court’s declaration of equality in public education and the ensuing focus on educational accountability is as valid and relevant today as they were when written 55 years ago.

In fact, Time magazine’s November 14, 2011 issue examined the United States of America’s current state of income inequality and its crippling dream of upward mobility. Foroohar’s (2011) article in this issue entitled “What Ever Happened To Upward Mobility?” concluded that the best hope in restoring the American dream is by improving education, which is “the factor that is more closely correlated with upward mobility than any other” (p. 34). A second article by Zakaria (2011) entitled “When Will We Learn?” went a step further by concluding that the best investment in education should be focused on teachers because they “produce the best results for students, more than class size or money or curriculum” (p. 44). As a poignant example, the same article estimated that if African American students were taught by a top-quartile teacher four years in a row
instead of a bottom-quartile teacher, it would be enough to close the black-white test-score gap.

Federal, state, and local governments are focused on improving education in order to expand opportunities for American children of all backgrounds while also ensuring that all children receive quality education. One critical approach towards accomplishing this goal is to keep teachers accountable while improving their practice and performance so that ultimately student achievement improves. Research in the past few decades demonstrates that teacher quality is one of the most important predictors of student achievement (Darling-Hammond, 1997b). Furthermore, a teacher’s effectiveness in teaching methods and their knowledge of the content area correlate to student achievement as well (Strahan, 2003). Thus, a teacher evaluation system holds the potential to play a vital role in effectively accessing teacher performance. With the use of multiple data sources and the incorporation of multiple evaluators, a multidimensional evaluation system can provide teachers with better feedback and assessment information to ensure teacher accountability in their performance and professional improvement. Peterson, Stevens, and Ponzio (1998) provided a concise rationale for the need of multiple data sources by writing, “no single data source works for all persons . . . because good teaching comes in a variety of forms and styles” (p. 124). This study will therefore help produce better understanding of how schools in Michigan that are already using research-supported teacher evaluation frameworks are moving toward developing multidimensional evaluation practices and systems.
CHAPTER II
LITERATURE REVIEW

Key Background Information

NCLB and Highly Qualified Teachers

With the authorization of the No Child Left Behind Act of 2001 (NCLB), federal focus has been increasingly placed on education by spotlighting student achievement and accountability (United States Department of Education, 2002). Student achievement became an important focus of this policy by mandating annual assessments for students and monitoring performance of specific subgroups through annual measurable objectives known as adequate yearly progress. In addition, schools were held accountable in their teacher hiring practices by requiring credentials for highly-qualified teachers (United States Department of Education, 2002). In Section 1119 of the bill, states and local school districts are instructed to “develop a plan to ensure that all teachers teaching in core academic subjects within the State are highly qualified” (United States Department of Education, 2002, p. 81). At a minimum, the legislation defines highly qualified teachers as holding a bachelor’s degree, full state certification or licensure, and a proven knowledge of the subject taught by typically passing a state exam. The premise is that highly qualified teachers will lead to enhanced quality of teaching and subsequently higher achievement in students. As Stronge and Tucker (2003) asserted, “Without capable, highly qualified teachers in America’s classrooms, no educational reform process can possibly succeed” (p. 3).

This movement towards establishing highly qualified teachers comes partially from research studies that correlated particular teacher characteristics with student
achievement. Some of these qualities known to impact student learning include teachers’ (a) verbal ability (Ehrenberg & Brewer, 1995), (b) content knowledge especially in mathematics and science (Goldhaber & Brewer, 1996), and (c) obtaining of a teaching license (Darling-Hammond, Berry, & Thoreson, 2001). An overarching study by Darling-Hammond (2000) consisted of a 50-state survey, case study analysis, and data from assessments in reading and mathematics administered by the National Assessment of Educational Progress to examine which school variables influence student achievement. She discovered that teacher knowledge and skills had statistically significant influences on student achievement; this variable was stronger than others were, such as teacher experience, class sizes, or pupil-teacher ratios. Her conclusion is similar to other studies conducted in this area that showed positive correlation between teacher qualification and student achievement (Strauss & Sawyer, 1986; Ferguson, 1991; Clotfelter, Ladd, & Vigdor, 2007; Croninger, Rice, Rathbun, & Nishio, 2007).

Though a highly qualified teacher is vital for improving education, it nonetheless does not ensure that teachers will produce greater student achievement. After all, qualification alone does not equal performance. For instance, with the NCLB federal mandate that all teachers should be highly qualified, most teachers now meet this requirement (Blank, Langesen, Laird, Toye, & de Mello, 2003). Nevertheless, overall student achievement scores have not increased with larger percentage of highly qualified teachers. Data from the National Assessment of Educational Progress (NAEP) indicated that Black and Hispanic students continue to lag behind White and Asian students in annual assessment data in reading and math (Lutkus, Rampey, & Donahue, 2005). Another report from the Harvard Center for Civil Rights examined math and reading
results according to race on the NAEP before and after the implementation of NCLB. Orfield (2006) summarized this study by concluding that under NCLB, “neither a significant rise in achievement, nor closure of the racial achievement gaps is being achieved” (p. 5). Several reasons besides teacher qualification are provided by researchers regarding NCLB’s deficiencies, ranging from unequal educational resources across schools (Darling-Hammond, 2006b) to questioning the reliability of standardized testing in assessing student learning (Hursh, 2007). Looking beyond these other factors for NCLB’s shortcomings, it is important to examine the role of teachers in achieving the goals of this legislation.

The NCLB legislation partially fell short in increasing student achievement because, as mentioned earlier, highly qualified teachers do not necessarily lead to highly qualified teaching. As an example, Pool, Ellett, Schiavone, and Carey-Lewis (2001) studied teachers who obtained a rigorous teaching certification through the National Board of Professional Teaching Standards (NBPTS) program. The goal of the program is to improve student learning by strengthening teaching standards and qualifications. The National Board Standards is not meant to replace state licensing but to complement it through advanced standards for experienced teachers. The case studies revealed that NBPTS certified teachers showed considerable variations in quality of teaching and student achievement across all grade levels. Other studies have also confirmed that teacher qualifications alone such as certification, experience, and advanced degrees do not categorically influence student learning (Goldhaber & Brewer, 1997; Goldhaber & Brewer, 2000). Tucker and Stronge (2005) made this clear when they said, “While
licensure or certification is a significant indicator of teacher quality, these factors alone are insufficient for teacher effectiveness” (p. 6).

**Impact of Teacher on Student Achievement**

The question remains as to what impacts student learning. It should first be mentioned that students must be held responsible for their share of learning. As Frymier (1998) wrote, “Because every person is accountable for his or her own behavior but not for what other people do, teachers must be held accountable for what they do as teachers but not for what their students do as learners. Students are responsible for their own learning” (p. 233). The author’s statement is valid to the extent that students should face positive or negative consequences for their own actions related to learning. Students’ role in learning should not be marginalized during this period of increased accountability in education. After all, without student participation, learning cannot happen. Then again, former Secretary of Education Richard W. Riley’s (1999) report pointed out that state standards and assessments are holding students responsible for their learning such that “students must earn promotion through effort and achievement” (p. 1).

Consequently, students are by no means absolved of their own responsibilities or seen as nonparticipants in the process of learning. Conversely, towards the end of his article, Frymier himself admitted, “Teachers must be held accountable for what they do as teachers” (p. 234). Teachers create the conditions where learning can take place. It is therefore important to note that learning cannot be the sole responsibility of the student; it is a partnership between student and teacher where both individuals are held responsible for fulfilling their roles.
Many years before the forming of the NCLB legislation, researchers and educators similarly debated on what variables influence student achievement. Some, like Frymier (1998), pointed to students as a main factor. Another study released by the U.S. Office of Education in 1966 known as the Coleman Report claimed that "schools bring little influence to bear upon a child's achievement that is independent of his background and general social context" (Coleman et al., 1966, p. 325). The study claimed that the variables of students’ background and social context were most important in influencing student achievement, which partially led to a perspective that teachers are absolved from responsibility for their students’ success or failure (Blankstein, 2004). However, a major flaw in the study was its focus on school-wide scores instead of individual teachers. As Marzano, Pickering, and Pollock (2005) made evident, “within a given school, though, there is a great deal of variation in the quality of instruction from teacher to teacher” (p. 3). With this perspective of examining individual teachers, numerous studies thereafter showed that educators have a strong influence on student achievement (Borman & Kimball, 2005; Hanushek et al., 2005; Nye, Konstantopoulos, & Hedges, 2004; Muijs, & Reynolds, 2002; Goldhaber, D. D., Brewer, D. J., & Anderson, D., 1999).

Some of these studies focused on specific school districts to find connections between individual teacher’s effects on their students’ achievement scores. For example, Haycock (1998) began her study of Boston Public Schools by stating that in order to increase the achievement levels of minority and low-income students, focus should be placed on good teaching. Data from the district showed that tenth grade students taught by the least effective teachers lost ground in math and made no gains in reading during one academic year. On the other hand, the top third of teachers produced up to six times
the learning growth compared to the bottom third of teachers. A second study by Aaronson, Barrow, and Sander (2002) examined Chicago school district for three years and found that “one semester with a teacher rated two standard deviations higher in quality could add 0.3 to 0.5 grade equivalents, or 25 to 45 percent of an average school year, to a student's math score performance” (p. 1). Lastly, a study conducted by Jordan, Mendro, and Weersinghe (1997) examined Dallas school district for four years. This quantitative study controlled for students’ effects of socioeconomic status, ethnicity, language proficiency, and gender based on outcome scores. Results showed that average sixth grade mathematics and reading scores were expected to increase from the 55th and 59th percentile to the 76th percentiles if students were assigned to three highly effective teachers in a row. Conversely, sixth grade mathematics and reading scores were expected to decrease from the 57th and 60th percentile to the 27th and 42nd percentile respectively if they were assigned to a series of three ineffective teachers during the same period. These studies essentially showed that student achievement varied within school districts according to individual teachers.

Besides research studies that examined specific school districts, other studies looked at data across various schools. For instance, Heck’s (2008) study included data of 9,196 students from 156 different elementary schools, examining successive teachers’ effects on student achievement. His results showed the effectiveness levels of successive teachers were positively associated with estimates of students’ ending achievement in both reading and math; in fact, over 80% of the variance in teacher effects was between classrooms. In another meta-analysis, Marzano’s (2003) research indicated that an effective teacher at an effective school can increase student achievement from 50th
percentile to the 96th percentile after two years. Borman and Kimball (2005) wrote that the difference between being taught by an effective teacher versus an ineffective teacher can lead to a full grade level of achievement. Another study by the University of Tennessee used statistical mixed-model methodologies to conduct multivariate, longitudinal analysis of student achievement (Wright, Horn, & Sanders, 1997). Results from this study showed that the effect of teachers is a dominant factor influencing student achievement gains; the authors of the study concluded, “More can be done to improve education by improving the effectiveness of teachers than by any other single factor” (Wright, Horn, & Sanders, 1997, p. 63).

Teacher Evaluation System

Subsequently, an important indicator for student growth is the individual teacher’s effectiveness in the classroom, which has a greater impact than teacher’s qualifications or other characteristics. Effective teaching will be defined later in this paper. A key question then follows: how is teacher effectiveness measured? Toch and Rothman (2008) wrote that teacher evaluation should be the center of the accountability portion of NCLB in order to measure effective teaching. After all, teacher evaluation matters because teaching matters. As Stronge and Tucker (2003) wrote, “Without high quality evaluation systems, we cannot know if we have high quality teachers” (p. 3). Teacher evaluation and teacher quality influence one another; the former helps in improving the latter (Schwartz, 2007). Tucker and Stronge (2005) used the analogy of a fitness trainer who improves his performance through guidance and evidence of effectiveness as seen through tangible results such as improved running time. Similarly, teacher evaluations provide feedback and guidance for improving professional practice. As research already
showed, quality teaching matters in that teachers influence what students learn. Therefore, by investing in improving teacher evaluation, student achievement is impacted in the end.

**Purpose and Objective of Teacher Evaluation**

Before examining the makeup of teacher evaluation systems, one must first provide the purpose for an evaluation system. Danielson & McGreal (2000) presented two primary objectives for evaluating teachers: professional growth and accountability. Beerens (2000) wrote that the purpose of teacher evaluation is to improve teacher effectiveness and encourage professional growth, and also to remediate or eliminate weak teachers. Peterson (2000) explained that the “most visible purpose of teacher evaluation is staffing decisions” while “the most discussed purpose of teacher evaluation is to improve practice” (p. 37). All these definitions hold in common two main dimensions involved in assessing teachers: summative and formative assessments.

Summative assessment is used to determine the future employment status of teachers. It fits into the accountability portion of a teacher evaluation system where teachers are held responsible in meeting determined performance standards. Therefore, this area of evaluation entails certain follow-up actions that Scriven (1967) listed as including selection of teachers to license, hire, give tenure, promote, demote or dismiss. In other words, the accountability portion reflects the need in determining competencies of teachers, which is also one of the major requirements in the NCLB legislation mentioned earlier. Finally, Danielson & McGreal (2000) wrote that a proper evaluation system encompasses the mission and goal of the individual school and district while at the same time serving and supporting individual teacher goals. With this compatibility,
the summative portion focuses specifically on measuring the performance of teachers, typically taking place after evidences are gathered and examined with due process. In summary, a summative evaluation involves making a judgment about teacher performance toward the end or summation of their activities. The focus is on outcome.

In contrast to the summative function of a teacher evaluation system, a formative assessment examines the performance improvement of teachers, reflecting the need for educators to continually grow and develop professionally. While the summative portion looks at the performance of teachers in accomplishing their individual and school-wide goal(s), the formative portion is focused on improvement. It takes place while the activities are in progress or forming. Formative assessment is important because as Stronge & Helm (1992) wrote, if established goals are to be achieved within a school, one should emphasize also the improvement and monitoring of progress towards those goals as well. This type of evaluation uses a supportive, nonjudgmental approach where teachers’ strengths and weaknesses are identified in order to further develop or establish a plan for support and training. The focus is on ongoing professional improvement.

Collinson (2001) provided succinct distinctions between summative and formative assessments respectively by writing, “One operates as a deficit model, the other as a growth model; one acts as the stick, the other as a carrot; one represents teacher passivity, the other, active teacher involvement; one is externally motivated, the other, internally motivated” (p. 151).

One group of researchers believes that summative and formative assessments should be separate from one another. For instance, Duke (1990) deemed that conventional teacher evaluation systems are prone to focus on the accountability or
summative portion, while omitting professional development; he therefore recommended “the complete separation of growth-oriented and accountability-based teacher evaluation” (p. 136). Popham (1988) further explained that without this separation, teachers will be reluctant to reveal their weaknesses or agree with their evaluator’s negative assessment since they are also liable to summative judgments. Lastly, Stroble and Cooper (1988) pointed to the problem for evaluator who will face a “role conflict” (p. 234) if summative and formative assessments are joined together since they must be both mentor and evaluator simultaneously. These authors in essence believe that judgment and assistance cannot properly coexist.

On the other hand, other researchers argue that summative and formative assessments should not be viewed as exclusive components within an evaluation system. After all, how is one able to effectively provide assistance without also making diagnostic judgments? As a comparison, teachers simultaneously serve as mentor and evaluator to their students. Stronge (1997) wrote, “Performance improvement and accountability purposes are not competing, but supportive interests—dual interests that are essential for improvement of educational service delivery” (p. 4). These dual interests are both essential because the formative evaluation identifies methods to improve performance while the summative evaluation determines whether the performance has improved sufficiently. Neither approach can serve a teacher and school properly on its own. Otherwise, when a summative evaluation takes place, a teacher will not be informed of areas of weaknesses that can and should be improved without the use of formative feedback. Similarly, formative evaluations without consequences of summative actions will not provide incentives for teachers to properly act on the feedback. Danielson &
McGreal (2000) further pointed out that summative assessment should be the summation of formative assessments. After all, “only through the collection of data over time . . . can an evaluator hope for a summative assessment to be fair and accurate” (Goldstein, 2005, p. 237).

In examining research studies that looked at the compatibility of summative and formative evaluations, Goldstein (2005) conducted a longitudinal mixed methods study of a district’s peer assistance and review program across 28 schools. Despite combining formative and summative functions with the evaluators in this program, teachers nonetheless reported a high degree of trust in their coaches. The few teachers who did not were all low-performing teachers; Goldstein argued that these teachers might express a negative relationship since they are the ones struggling. Another study by Milanowski (2005) compared two groups of teacher within a large urban district. One group had split the formative and summative roles between two evaluators while the other group had both roles carried out by the same person. Results showed “no major differences between the groups in terms of openness to discussion of difficulties, reception and acceptance of performance feedback, stress, turnover intentions, actual turnover, or performance improvement” (Milanowski, 2005, p. 153). On the contrary, the author found that few teachers in the split group made use of the developmental exercises to help them improve on the summative evaluation. Furthermore, teachers in the split group expressed negative reactions to their summative evaluators due to less communication opportunities and the perceived unfair practice of being assessed but not given any developmental assistance. Milanowski therefore suggested that formative and summative processes should be coordinated at minimum.
It should be mentioned here that other more important factors should be considered when examining the formative and summative functions of evaluation. As Milanowski (2005) remarked, “the quality of developmental assistance, the credibility and accessibility of the mentor or mentor/evaluator, and the personality compatibility of the evaluator and evaluatee, are more important for the effectiveness of evaluation” (p. 165). Additionally, going back to Goldstein’s (2005) study, trust factor between teacher and evaluator plays an instrumental role as well. In fact, with greater collaboration of summative and formative evaluations, the issues of trust, credibility, and accessibility are more likely mitigated. In summary, evaluation systems that make use of both summative and formative functions are essential and appropriate in order for evaluations to constructively accomplish both the needs of the individual teacher and the requirements from their school.

Deficiencies in Traditional Teacher Evaluation Systems

Even though a teacher evaluation system should fulfill the two primary objectives of teacher accountability and professional growth, traditional teacher evaluation systems have unfortunately fallen short in these objectives. Frase and Streshly (1994) summarized the state of teacher evaluation when they wrote, “Research and learned opinion strongly support the contention that teacher evaluation has been of little value” (p. 48). There are several reasons for the ineffectiveness and inadequacy of traditional teacher evaluations systems. First of all, the summative objective of an evaluation system is not met. Teachers are not properly held accountable in fulfilling their professional duties. Brandt et al. (2007) reported that tenured teachers’ evaluation only took place once every three to five years with as little as one class period. The limited
time period for evaluation can lead to limited feedback by evaluators. Another reason for
the lack of accountability is deficiencies in professional standards within traditional
evaluations systems. The system is seen as a checklist-oriented process with a lack of
clear definitions of job roles, functions, and expectations (O’Day, 2002). As an example,
these checklist items are simplified to measurable categories that are not central to
instruction, such as neatness of classroom bulletin boards (Donaldson, 2010). Another
study by Brandt et al. (2007) examined teacher evaluation policies across school districts
in the Midwest region and found that “district policy documents were more apt to specify
the processes involved in teacher evaluation (who conducts the evaluation, when, and
how often) than they were to provide guidance for the content of the evaluation, the
standards by which the evaluation would be conducted, or the use of the evaluation
results” (p. i-ii). In another study that looked at five urban school district’s evaluation
systems, though the researchers found common set of standards, they failed to find
documentations defining neither those standards nor any performance rubrics (Koppich &
Showalter, 2008).

Because of these deficiencies in clear guidelines and standards and limited
number of evaluations, evaluators are not able to properly hold teacher accountable. It
should consequently not be surprising that traditional teacher evaluations usually have
inflated scores. For example, Tucker (1997) conducted a study across 134 public schools
in Virginia that found principals with a staff of 100 teachers identified less than 2% as
incompetent tenured teachers, while remediating 0.68% teachers and recommended
dismissal for 0.10% teachers. Secondly, an investigative study by Scott Reeder (2005)
found that in the past decade, 83% of Illinois school districts have never rated a tenured
teacher as unsatisfactory. Finally, a study by Weisberg et al. (2009) conducted in 12
school districts in Arkansas, Colorado, Illinois, and Ohio found that less than one percent of surveyed teachers received a negative rating on their most recent evaluations. Though these low percentages can be seen as validating overall superior teacher performance, in actuality, they do not correlate with research studies that estimate that 5 to 15% of teachers are marginal or incompetent (Tucker, 1997). In fact, Weisberg et al. (2009) pointed out that “over the last three years, only 10 percent of failing schools issued at least one unsatisfactory rating to a tenured teacher” (p. 12). Donaldson’s (2010) report goes on to admit, “Both principals and teachers believe that teachers are less effective than ratings indicate” (p. 55). Finally, Frase and Streshly (1994) concluded their section on “inflation of teacher evaluation ratings” by stating, “A plethora of studies and observations reveal evaluation ratings are grossly inflated” (p. 48). Traditional teacher evaluation systems have fallen short in keeping teacher accountable.

In addition to the summative portion of traditional teacher evaluation system falling short in fulfilling its purpose, the formative function that is focused on performance improvement has typically failed as well. One reason for this neglect is the lack of standards within the evaluation system (Brandt et al., 2007; Koppich & Showalter, 2008). Without established standards for effective instruction, evaluators are not able to provide suggestions for teacher improvement. Furthermore, many of the deficiencies in the formative portion are connected to problems in the summative section already mentioned above. As schools are not able to either hold teacher accountable or dismiss consistently poor performing teacher, it leads to both the lack of focus on growth and failure in recognizing excellence among top-performers. In effect, all teachers are seen as equal performers (Weisberg, Sexton, Mulhern, & Keeling, 2009). Thirdly, Duke
(1990) identified that conventional teacher evaluation systems tend to exclude focus on professional development. In another study, Weisberg et al. (2009) wrote that after studying 12 districts, only 26% of teachers during their most recent evaluations were identified with areas of development; this means that 3 out of 4 teachers were evaluated without feedback on ways to improve practice. Duffett, Farkas, Rotherham, and Silva’s (2008) survey of K-12 public school teachers found that only 26% of teachers responded that their evaluation “was useful and effective in terms of helping you be a better teacher” (p. 3). Weisberg et al. (2009) summarized the current state of traditional teacher evaluation systems accordingly: “Excellent teachers cannot be recognized or rewarded, chronically low-performing teachers languish, and the wide majority of teachers performing at moderate levels do not get the differentiated support and development they need to improve as professionals” (p. 6). Clearly, traditional teacher evaluation systems have failed in improving and developing teachers.

In summary, traditional teacher evaluation systems have fallen short in providing summative and formative evaluations. In fact, Loup, Garland, Ellett, and Rugutt (1996) showed that teacher evaluation practices in the 100 largest school districts in the United States showed little change during the past 10 to 15 years despite increased attention in educational accountability and school reform. With a lack of proper evaluation systems for teachers, student achievement is impacted. As Holland (2006) wrote, traditional evaluation systems have shown little to no effect on improving accountability and student achievement. This leads to traditional teacher evaluations falling short in adequately assessing the multifaceted and broad scope of the position, which brings about failures in producing a positive impact on professional growth (Danielson & McGreal, 2000).
What's more, the American Federation of Teachers (2011) began its press report regarding the current state of teacher evaluation accordingly: “With rare exceptions, teacher evaluation procedures are broken—cursory, perfunctory, superficial and inconsistent” (p. 1). Unfortunately, these negative factors in the end result in lack of both professional growth and student achievement.

Focus of the Study

Multidimensional Teacher Evaluation System

Researchers have argued that a well-designed teacher evaluation system should be an important component of school improvement (Normore, 2005). Unlike traditional evaluation systems, a well-designed teacher evaluation system should be “multidimensional” (Yonghong & Chongde, 2006, p. 29) with “the use of multiple measures” (Darling-Hammond, 2006a, p. 135) where “individual teachers in a system are evaluated with different kinds of data” (Peterson, 2004, p. 63) through the incorporation of “multiple assessors” (Odden, 2004). This type of evaluation system is needed because the art and science of teaching is complex. As Danielson (1996) made evident, “a teacher makes over 3,000 nontrivial decisions daily” and the tasks of teaching are “similar to not one but several other professions” (p. 2). As a result, for any given individual teacher, no single data source is valid or feasible; rather, multiple and variable data sources are needed to accurately and fairly assess teachers (Peterson, 2006). Stronge (2006) further asserted, “As multiple data sources are properly employed in performance evaluation, the validity and utility of the process can be dramatically enhanced” (p. 11). Moreover, evaluators are strengthened by additional data sources that support their assessment. Peterson (2006) specifically pointed out that a multidimensional teacher evaluation
system can take pressure off administrators, who are no longer the single data source. Finally, a multidimensional teacher evaluation system builds a constructive environment of collaboration since multiple data is gathered from multiple stakeholders such as client surveys that engage students and parents, peer reviews and teacher portfolios that involve teachers, and pupil achievement results that provide less biased data.

With the current movement in research away from traditional to a multidimensional teacher evaluation system, there is already ample research on the shortcoming of the traditional teacher evaluation system that was mentioned earlier. In fact, research article titles such as Charlotte Danielson’s “New Trends in Teacher Evaluation” (2001) and Linda Darling-Hammond’s “Assessing Teacher Education: The Usefulness of Multiple Measures for Assessing Program Outcomes” (2006a) combined with latest headings such as “Teacher Evaluation 2.0” (New Teacher Project, 2011) point to current movement towards developing a multidimensional teacher evaluation model. As Peterson (2004) summarized, “researchers conclude that multiple data sources should be used in teacher evaluation” (p. 63).

Studies in various school districts have shown many advantages of using a multidimensional teacher evaluation system. For instance, as far back as 1987, Peterson’s study using descriptive and correlative analysis of a Utah school district’s evaluation results discovered that by using multiple data sources, different constructs of teacher quality were seen. The district utilized data sources that included administrator report, student and parent surveys, teacher tests, peer review, and teacher’s professional activities and years of experience. Results showed that the lines of evidence from these data sources were generally independent of one another with the mean absolute
correlation between data sources of 0.15. Peterson (1987) thus concluded, “This finding suggests that the multiple lines of evidence may have assessed different constructs of quality” (p. 316). Another study by McLarty (1985) examined the Tennessee Career Ladder Evaluation system, which used a multiple data source approach to assessment. The evaluation data came from four different evaluators: the teacher candidate, the school principal, three peer teachers, and twenty or more students. Data sources included instruments such as classroom observations, questionnaires for students and peer teachers, portfolio ratings, and written tests. When the study analyzed each of these instruments separately, it produced low intercorrelations, which provided initial evidence that each source measured different viewpoints or aspects of teacher performance. In a later report, Heneman, Milanowski, Kimball, and Odden (2006) conducted a multiyear mixed method study to investigate the validity of different teacher evaluation systems in four districts. Though the study found variability between teacher evaluation scores and student achievement gains across the four sites, Vaughn Schools and Cincinnati District showed similar and higher average correlation during the three years of study with 0.37 and 0.32 in reading and 0.26 and 0.32 in mathematics respectively. The authors noted that correlations ranging from 0.20 to 0.40 are considered meaningful indicators of an association between variables, which the other two districts did not display. The researchers therefore posited that the higher correlation results at Vaughn Schools and Cincinnati District was due partly because these districts utilized a greater number of data sources within their evaluation system. A different study by Den Brok, Van Eerde, and Hajer (2010) reviewed required competencies for teaching in multicultural schools through case studies. At the conclusion of the study, the researchers specifically noted
the following: “It was shown that multiple data collection techniques are needed to meet
the requirement of specificity and variability of competencies. This more comprehensive
nature allows for teachers to make their interactive knowledge explicit and to show a
larger part of their repertoire” (Den Brok, Van Eerde, Hajer, 2010, p. 729). Lastly, a
recent large scale study by MET project (2012) examined the practice of 1,333 teachers
across six districts who taught math and English language arts in grades four to eight.
The trained project raters scored 7,491 videos of lessons at least three times using
different research-supported observations rubrics. State tests scores, supplemental tests,
and student surveys were also incorporated from more than 44,500 students. The
quantitative study revealed that combining observations scores with evidence of student
achievement gains and student surveys improved predictive power and reliability.
Specifically, “depending on the classroom observation instrument used, the reliability of
the combined measure ranged from 0.55 to 0.67 in math and 0.51 to 0.54 in ELA. In
other words, combining multiple measures led not only to higher predictive power but
greater reliability (more stability) as well” (MET, 2012a, p. 10).

As a summary of the many benefits of a multidimensional teacher evaluation
system, Stronge (2006) provided the following elements from his synthesis of various
studies:

- Produce a richer textured and more complete portrait of the evaluatee’s
  performance
- Collect data in more naturally occurring situations
- Integrate primary and secondary data sources in the evaluation
- Assure greater reliability in documenting performance
• Enhance objectivity in documenting performance
• Document performance that is more closely related to actual work
• Offer a more legally defensible basis for evaluation decisions (p. 10).

Multiple data sources with multiple evaluators make for a more complete and more comprehensive teacher evaluation system, which is needed for evaluating the complex task of teaching.

Federal and State of Michigan Policy

As research is focusing on multidimensional teacher evaluation systems, policy makers are also beginning to emphasize the importance of research-supported evaluation instruments and frameworks. Starting at the federal level, following NCLB, the American Recovery and Reinvestment Act (ARRA) of 2009 targeted a significant amount of funding to improve state and local education systems. In particular, the Race to the Top Fund provided $4.35 billion in competitive grants for states; these grants are based upon meeting certain criteria, one of which specifies that states should “design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that differentiate effectiveness using multiple rating categories” (United States Department of Education, 2009, p. 9). Soon after this funding, Learning Point Associates (2010) published a report examining emerging trends in the Race to the Top grant application; the report concluded that a “common theme in the state legislation is the reworking of teaching standards . . . as specific steps in reforming teacher evaluation” (p. 12).

Before examining Michigan state policies, it is first important to understand the state’s current state in public K-12 education. The Education Trust-Midwest’s January
2011 report discovered that Michigan’s performance in both mathematics and reading fell across time relative to other states in the National Assessment of Educational Progress. As examples, over the last six years the state’s rank in fourth-grade reading dropped from tied for 25th place to tied for 34th, and in eighth-grade math, Michigan dropped from 34th place to tied for 36th. More specifically, the national math and reading results indicated that only 31 and 30 percent of Michigan students rank as proficient respectively. The same report emphasized that these lower rankings hold true for all students; “Group for group—poor, rich, white, black and Latino—our children perform well below their counterparts in other states” (p. 4). Secondly, the Schott Foundation for Public Education (2009) published another report that provided a state-by-state comparison of public education’s academic proficiency and access, which helped determine a metric for student’s opportunity to learn. Michigan ranked at the bottom of providing both low-proficiency and low-access public education to their students. Thus, Michigan has clearly fallen behind in K-12 public education. On the other hand, it is important to note that both of these reports went on to identify highly prepared and effective teachers as a core building block for successful improvement efforts. In fact, the Education Trust-Midwest report (2011) specifically wrote that Michigan is now beginning to look at “what must be done to make up for decades of neglect in areas like teacher evaluation” (p. 6).

It is therefore worth noting that Michigan has recently taken steps forward in establishing state policies focused on teacher performance. For instance, school reform laws were passed in 2009 in support of Michigan’s original Race to the Top application. The law, listed as Public Act 205, includes sections 1249 and 1250, known as “Race to
the Top” School Reform, which required districts to conduct annual educator evaluations that include student growth as a significant factor (S. Res. 981). A later memo from the Michigan Department of Education clarified the word “significant” by defining it as 40-60% of the overall evaluation (S. Vaughn, memo, February 8, 2010). Most recently, the governor of Michigan signed House Bill 4627 on July 19, 2011 which specified that eventually in the 2015-2016 school year, at least 50% of the annual year-end evaluation shall be based on student growth and assessment data. Michigan Department of Education distributed $1.3 billion to school districts through the federal State Fiscal Stabilization Fund (SFSF) towards the implementation of several activities, including educator evaluations. Specifically, these Michigan school reform laws required districts to do the following (Olivares, 2011):

- Conduct annual educator evaluations.
- Include measures of student growth as a significant factor.
- Locally determine the details of the educator evaluations, the consequences, and the timeline for implementation.
- Tie educator effectiveness labels to decisions regarding promotion and retention of teachers and administrators, including tenure and certification decisions.
- Use a performance-based compensation method that evaluates performance based, at least in part, on student growth data.

Michigan Senate Bill 1509 provided an effective date of September 1, 2011 for the new evaluation systems to begin for teachers and administrators (S. Res. 1509). It is worth noting that districts are encouraged by the Michigan guidelines for education evaluation system to include multiple data as measures of educator effectiveness. The reason for
doing so is explicitly stated as: “Educator evaluations are more appropriate and provide a
gericher source of information to the educators and to other stakeholders if they include
multiple measurements of educator effectiveness” (Olivares, 2011). In summary, federal
and state policy makers are joining with researchers in the current movement towards
emphasizing the development and implementation of multidimensional teacher
evaluation systems.

Effective Teaching

An important question thereby follows from this trend towards establishing
multidimensional research-supported evaluation instruments and frameworks: How can
teacher effectiveness properly be measured within a teacher evaluation system?

Definition of teacher quality and effective teaching. Before looking at specific
frameworks for measuring teacher effectiveness, one must first have a common
understanding of the qualities involved in effective teaching. In other words, what is
teacher quality? It must first be mentioned that teaching is a complex activity that
involves multiple variables such as behaviors, skills, and competencies (Darling-
the following model shown in Figure 1 to break down the components of teacher quality
into three main areas with explanation for each (p. 155):
First of all, teacher qualification has an impact on teacher quality. Like any other profession, teachers should have certain knowledge and skills that are needed for their practice. Researchers have shown a positive correlation between teacher qualification and student achievement (Strauss & Sawyer, 1986; Ferguson, 1991; Clotfelter, Ladd, & Vigdor, 2007; Croninger, Rice, Rathbun, & Nishio, 2007). This is why No Child Left Behind (2002) legislation required teachers to be “highly qualified.” At minimum, the legislation defines highly qualified teachers as holding a bachelor’s degree, full state certification or licensure, and a proven knowledge of the subject taught by typically passing a state exam (United States Department of Education, 2002). This first component of teacher quality is mainly addressed in the hiring and interview process. But as mentioned earlier, though a highly qualified teacher is vital for improving education, it nonetheless does not ensure that teachers will produce greater student achievement. In other words, it is necessary but not sufficient. After all, qualification alone does not equal performance.

Figure 1. Components of teacher quality.
Tucker and Stronge (2005) indicated in the model above that the second and third components of teacher quality include the much more complex activity of teaching. Instead of focusing on certain attributes and traits of teachers, these final two components look at what teachers do; the practices and outcomes associated with effective teaching. The second component, listed as teacher behaviors, examines the act of teaching while the last component, labeled teacher outcomes, focuses on the results of this teaching. It is these two areas that will be the focus of this study on teacher evaluation.

The question still remains on how to define the practice of effective teachers. Darling-Hammond (1998) provided an overall perspective of effective teachers when she wrote, “Teachers learn best by studying, doing, and reflecting; by collaborating with other teachers; by looking closely at students and their work; and by sharing what they see” (p. 6). All these behaviors help in defining an effective teacher. The National Board for Professional Teaching Standards (1989) went a step further by developing the following five core propositions to assist in the identification and recognition of effective teachers:

1. Teachers are committed to students and learning.
2. Teachers know subjects they teach and how to teach subjects to students.
3. Teachers are responsible for managing and monitoring student learning.
4. Teachers think systematically about their practice and learn from reflection.
5. Teachers are members of learning communities.

These two sources provide a sound agreement on the practices of effective teachers that enhance student learning. What's more, Marzano, Pickering, and Pollock (2005) have actually quantified the average effects of specific instructional strategies. For instance,
their research showed that instructional strategies such as summarizing and note taking and identifying similarities and differences can potentially lead to student achievement percentile gains of between 29-45 points. In essence, the aforementioned studies provide a common understanding of the qualities for effective teaching.

**Danielson’s Framework for Teaching.** In describing the qualities of effective teachers, it is then important to establish these qualities into a framework within a multidimensional teacher evaluation system. One of these research-informed standard and framework that emerged in the 1990s is Charlotte Danielson’s (1996) *Framework for Teaching*. The framework “identifies those aspects of a teacher’s responsibilities that have been documented through empirical studies and theoretical research as promoting improved student learning” (Danielson, 1996, p. 1). In doing so, the framework also helps to answer the question of what effective teachers know and do. Such a framework is beneficial for teachers at all levels, from new teachers to master teachers who are describing their practice to others.

Before examining the Danielson framework, it is first important to explain the research behind its formation. Danielson (2007) explained that her framework is “based on the Praxis III criteria developed by the Educational Testing Service (ETS) after extensive surveys of the research literature, consultation with expert practitioners and researchers, wide-ranging job analyses, summaries of the demands of state licensing programs, and fieldwork” (p. 183). The Praxis series was originally designed for use by state or local agencies when making teaching licensing decisions. In fact, approximately 40 states and U.S. territories currently require the Praxis Series tests for teacher licensure (Educational, 2012). Praxis III was later developed for assessing beginning teachers’
skills and performance. The knowledge base of the Praxis Series was gathered through seven years of research, from 1987 to 1993, that consisted of job analysis of elementary, middle, and high school teachers combined with extensive literature search on effective teaching. Findings were then subjected to pilot and field testing by experts with assistance from the American Association of Colleges for Teacher Education, the American Federation of Teachers, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the National Association of State Directors of Teacher Education and Certification, and the National Education Association (Danielson, 2007; Myford et al., 1993). Essentially, the Danielson framework draws from the same research base as the criterion for the Praxis Series.

It is also important to examine the major assumptions behind the Danielson framework. Danielson (2007) stated that her framework is based on the learning assumption that it is important for all students “to acquire deep and flexible understanding of complex content, to be able to formulate and test hypotheses, to analyze information, and to be able to relate one part of their learning to another” (p. 15). Danielson then proceeded to write that this type of learning is acquired through a constructivist approach that recognizes that it is the learner who does the learning; in other words, the students’ understanding of content depends on their experience in obtaining it themselves. The teacher’s role is to guide their students towards understanding that content.

Based on these fundamental assumptions, the Danielson framework is organized according to four domains or realms of activity that refer to distinct aspects of teaching. These domains come from the aforementioned Praxis series. Furthermore, these domains
align with the Interstate New Teachers Assessment and Support Consortium (INTASC) standards, which represent the professional consensus of what a beginning teacher should know” (MET, 2010, p. 2). Danielson (2007) identified the four domains as planning and preparation, the classroom environment, instruction, and professional responsibilities. The first domain of planning and preparation examines how the teacher organizes the content that students will learn; in other words, it focuses on the instructional design. Emphasis is placed on teachers grasping rich knowledge of the content and outlining instruction for diverse classroom. The second domain of classroom design moves from content to the examination of the classroom environment. This domain is based upon the understanding that the classroom environment is a vital facet of a teacher’s ability in encouraging a culture of learning that makes their practice of instructional skills possible; therefore, emphasis is placed on teachers creating an environment that is conducive to learning. The third domain of instruction captures the essence or heart of teaching, where teachers bring content to life for their students; it is the implementation of the first planning domain. Emphasis in the third domain is placed on teachers’ ability to engage students in learning the content. Finally, the last domain of professional responsibilities centers on the role of educators outside of their classroom. It places an emphasis on teachers demonstrating a commitment to high ethical and professional standards with a focus on continual professional growth. This last domain encompasses a wide range of responsibilities such as self-reflection and professional growth to interaction with families of students. Danielson lists a number of components for each of the four domains as seen in Table 1:
### Table 1

*Framework for Teaching*

<table>
<thead>
<tr>
<th>Domain</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1: Planning and Preparing</td>
<td>1a: Demonstrating knowledge of content and pedagogy</td>
</tr>
<tr>
<td></td>
<td>1b: Demonstrating knowledge of student</td>
</tr>
<tr>
<td></td>
<td>1c: Selecting instructional goals</td>
</tr>
<tr>
<td></td>
<td>1d: Demonstrating knowledge of resources</td>
</tr>
<tr>
<td></td>
<td>1e: Designing coherent instruction</td>
</tr>
<tr>
<td></td>
<td>1f: Assessing student learning</td>
</tr>
<tr>
<td>Domain 2: The Classroom Environment</td>
<td>2a: Creating an environment of respect and rapport</td>
</tr>
<tr>
<td></td>
<td>2b: Establishing a culture of learning</td>
</tr>
<tr>
<td></td>
<td>2c: Managing classroom procedures</td>
</tr>
<tr>
<td></td>
<td>2d: Managing student behavior</td>
</tr>
<tr>
<td></td>
<td>2e: Organizing physical space</td>
</tr>
<tr>
<td>Domain 3: Instruction</td>
<td>3a: Communicating with students</td>
</tr>
<tr>
<td></td>
<td>3b: Using questioning and discussion techniques</td>
</tr>
<tr>
<td></td>
<td>3c: Engaging students in learning</td>
</tr>
<tr>
<td></td>
<td>3d: Using assessment in instruction</td>
</tr>
<tr>
<td></td>
<td>3e: Demonstrating flexibility and responsiveness</td>
</tr>
<tr>
<td>Domain 4: Professional Responsibilities</td>
<td>4a: Reflecting on teaching</td>
</tr>
<tr>
<td></td>
<td>4b: Maintaining accurate records</td>
</tr>
<tr>
<td></td>
<td>4c: Communicating with families</td>
</tr>
<tr>
<td></td>
<td>4d: Participating in a professional community</td>
</tr>
<tr>
<td></td>
<td>4e: Growing and developing professionally</td>
</tr>
<tr>
<td></td>
<td>4f: Showing professionalism</td>
</tr>
</tbody>
</table>

Essentially, the four domains of teaching contain a total of 22 components to identify a range of appropriate teacher behaviors. Each of these elements includes rubrics to assess unsatisfactory, basic, proficient, and distinguished performance.

It should be mentioned here that these components and elements of teaching are meant to serve as a structure within which teachers can situate their actions. It is not
meant to be checklists of specific behaviors since specific actions vary in teaching depending upon the particular teacher and context. In other words, though an effective teacher exhibits practices listed in the table above, they do not achieve them in the same way because no particular approach will be effective in every situation. As an example, Component 3a of communicating with students represents the effect achieved, while the framework avoids listing specific actions taken to provide such communication. The express action is the choice and decision that professional teachers make depending on their own situation and context, which they justify to their evaluator. Thus, the purpose of the Danielson framework is to provide common definitions of expertise and procedures for effective teaching. As Danielson (1996) explained, “Because teaching is complex, it is helpful to have a road map through the territory, structured around a shared understanding of teaching” (p. 2). Such a framework or roadmap will also serve a vital function in evaluating teacher performance.

A number of studies have shown the efficacy of using the Danielson framework within a teacher evaluation system (Heneman, Milanowski, Kimball, & Odden, 2006; Holtzapple, 2003; Kimball, White, Milanowski, & Borman, 2004; MET, 2012; Sartain, Stoelinga, & Krone, 2010; Taylor & Taylor, 2003; Toch & Rothman, 2008). Districts in many states such as Chicago, Cincinnati, and Las Vegas have embraced the Danielson framework as their performance evaluation rubric; hence, there is a growing body of research related to the use of this model. For instance, the Teacher Advancement Program, a widely implemented leadership development system for teachers, uses Danielson’s Framework for defining the teaching competencies at each level of teacher performance (Toch, 2008). Furthermore, researchers from the Consortium on Chicago
School Research at the University of Chicago published a report in late 2010 that stated, “the Charlotte Danielson Framework for Teaching, which attempts to delineate the observable components of effective teaching, is perhaps the most well-known example” of a tool that measure teacher effectiveness (Sartain, Stoelinga, & Krone, p. 4). The report went on to examine Chicago Public Schools using the Danielson framework and concluded that the instrument is a reliable tool for measuring components of teaching (Sartain, Stoelinga, & Krone, 2010). With the use of multi-facet Rasch analysis, Sartain, Stoelinga, and Brown (2009) specifically cited in their analysis of 277 classroom observation data from Chicago Public Schools’ evaluation system that the reliability for finding strong and weak teachers using the Danielson framework was high (reliability=0.94, separation=3.92), where reliability meant “teachers with estimated measures of high teaching ability actually were more successful in the classroom during the observed lesson than teachers with estimated measures of low teaching ability” (p. 24). The same study produced qualitative data through interviews of 39 principals and 26 teachers which showed that 84% of principals and 100% of teachers held mostly positive attitudes or mixed perceptions about the framework as opposed to mostly negative attitudes. These positive comments centered on improvements in the following major areas: principals/teachers were better prepared for conferences and lessons, clearer expectations, improved relationships, more reflection discussions, and improved focus on instruction.

Another study at a large Western school district conducted an analysis of the relationship between scores of a standard-based teacher evaluations system modeled on Danielson’s framework and student achievement measures. The study “provided some
initial evidence of a positive association between teacher performance, as measured by the evaluation system, and student achievement” (Kimball, White, Milanowski, & Borman, 2004, p. 2). A similar study by Kane, Taylor, Tyler, and Wotten (2011) examined Cincinnati Public Schools’ Teacher Evaluation System (TES), based on the Danielson framework, with the primary goal of examining the relationship between teachers’ evaluation rating and their assigned students’ test-score growth. Results after analyzing 365 teachers in reading and 200 teachers in math showed “that teachers’ classroom practices, as measured by TES scores, do predict differences in student achievement growth” (Kane, Taylor, Tyler, and Wotten, 2011, p. 58). Specifically, their data showed that a score improvement by one point in a teacher’s overall classroom practice is associated with one-seventh of a standard deviation increase in reading, and one-tenth of a standard deviation increase in math. Furthermore, when these researchers analyzed teachers who were evaluated by TES multiple times in different years, they discovered that these teachers’ improvements in practice over time, as measured by TES, correlated to improvements in their students’ achievement gains. Finally, a more comprehensive study by Heneman, Milanowski, Kimball, and Odden (2006) examined the relationship between teacher evaluation scores of systems that utilized the Danielson framework with actual achievement of their students. The researchers correlated evaluation scores with estimates of the value-added academic achievement in math and reading of their students, which were further controlled for prior achievement and other student characteristics that influence student learning such as socioeconomic status. Plus, three years of data were included for further reliability. Ultimately, the study found a positive relationship between teacher evaluation scores and student achievement with one
district showing an average correlation over three years of 0.37 in reading and 0.26 in mathematics while another district showed similar correlations with 0.35 in reading and 0.32 in mathematics, which are “both considered meaningful indicators of an association between variables” in educational research (Heneman, Milanowski, Kimball, & Odden, 2006, p. 4). Furthermore, when these researchers followed up these quantitative results with teacher and administrator feedback through interviews and surveys, they discovered that in general, educators accepted the model as suitable and as sufficient description of teaching; in fact, most teachers perceived the ratings as accurate and fair.

A recent study by the Measures of Effective Teaching Project (2012a) funded by the Bill and Melinda Gates Foundation examined the Danielson framework by studying the association between the framework and student outcome. This large-scale study is based on the practice of 1,333 teachers from six districts who taught math or English language arts in grades 4 through 8. The trained raters scored 7,491 videos of lessons using the Danielson framework. The study also incorporated state test scores from more than 44,500 students; these scores were examined based on achievement gains, which “is the average difference between students’ actual and expected achievement test score at the end of the year across all tested students in a classroom who have a prior year achievement test score” (MET, 2012a, p. 42). To answer potential concerns about the quality of the particular state tests, the study also utilized two other supplemental assessments proven to have high levels of reliability across different groups of students: the Balanced Assessment in Mathematics (BAM) and the open-ended version of the Stanford 9 reading test. Results showed that the Danielson framework observation instrument is positively associated with student achievement gains in both subjects across
the state test and the supplemental assessment. Specifically, the relationship between teacher observation scores and state math and English language arts achievement gains showed correlation of 0.18 and 0.11 respectively. It should be noted that these positive correlations were seen despite the study not incorporating two of the four components due to limitations of only using videos of classroom practice to access teacher performance. Nonetheless, the study concluded from its results that “teachers who more effectively demonstrated the types of practices emphasized in the instrument had greater student achievement gains than other teachers” (MET, 2012b, p. 5). In other words, since the study showed alignment between observation scores using the Danielson framework and student achievement gains, it provides further evidence that the framework actually identifies instructional practices shown to improve student learning.

On the other hand, some researchers have pointed to a smaller number of studies that have shown mixed results when accessing evaluation systems using the Danielson framework. For instance, Borman and Kimball (2005) found that teachers in a Nevada school district who received higher scores using the Danielson standards improved in closing the achievement gap in reading but showed little progress in math. Another study by Gallagher (2004) discovered that an elementary charter school in Los Angeles Unified School District displayed a positive and statistically significant relationship between teacher evaluation scores that used the Danielson framework and student achievement in reading, but in math, this relationship lacked statistical significance. These types of studies show the need for examining other factors that influence research findings such as variations in evaluator training, the degree of adherence to the Danielson framework, and the number and type of instruments used. For instance, Borman and Kimball’s study
admitted that the district made changes to the Danielson evaluation rubric. The second study by Gallagher stated that the variation of results was partially due to teachers and evaluators having a more pedagogical knowledge and better alignment to standards and assessment in reading than in math. In essence, these other factors help to partially explain the variation of results in these studies. It should be mentioned that previously cited studies that showed positive correlations between teacher evaluations using the Danielson framework and student achievement also attest to the importance of considering factors outside the framework itself. For instance, Heneman, Milanowski, Kimball, and Odden (2006) wrote that higher average correlations were seen in districts that provided intensive, high-quality training and also utilized multiple evaluators. The qualitative portion of Sartain, Stoelinga, and Brown’s study (2009) pointed out that principals and teachers likewise admitted to the need for more training on the framework; they also expressed concerns about limited time to conduct evaluations. It is therefore valuable and worthwhile to address these other factors outside the actual framework itself when implementing a teacher evaluation system.

In addition to the aforementioned studies showing the efficacy of using the Danielson framework, various states and districts have also incorporated the Danielson framework into their evaluation system. At the national level, a recent report by Licciardi (2010) found that a number of the Race to the Top applicant states are utilizing the Danielson framework to guide the development of their evaluation process. Furthermore, Learning Point Associates’ (2010) review of Race to the Top applicants found that out of the 16 states that identified specific frameworks they are using for developing their evaluation rubrics, half of them were using the Danielson framework. In fact, the
following title of a recent news article attests to the recognition and reputation of this framework: “Charlotte Danielson's Framework for Teaching Sees Record Growth as States and Districts Redefine Teacher Evaluation” (Charlotte, 2011). The news report goes on to explain that besides numerous statewide adoptions, the Danielson framework is used by “hundreds of districts in other states” that include Los Angeles Unified School District, Pittsburgh Public Schools, Hillsborough County Public Schools and 14 other districts throughout the state of Florida (Charlotte, 2011). As far as Michigan, the state’s Department of Education (2011) website lists the Danielson framework as a key resource tool to help inform local educator evaluation efforts. Consequently, even though there are other more recently published research-supported frameworks and instruments for teacher evaluation such as McREL’s Teacher Evaluation System, Classroom Assessment Scoring System, and statewide evaluation systems such as Tennessee’s Framework for Evaluation & Professional Growth and Denver’s ProComp, Danielson’s framework nonetheless is one of the most widely used with one of the greatest longevity in the field (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007; Charlotte, 2011). Ultimately, the framework offers a valuable starting point and roadmap for a teacher evaluation system by providing a common understanding of the qualities of effective teaching.

**Components of Multidimensional Teacher Evaluation System**

The Danielson framework provides a common definition of effective teaching. A multidimensional teacher evaluation system must also make use of instruments to collect data and access teachers according to the framework. The following section includes a listing of research-proven components and instruments used within a teacher evaluation system. These instruments include classroom observations by principals, student
achievement data, teacher portfolio, and peer and student evaluation of teachers (Danielson & McGreal, 2000; Peterson, 2000; Stronge, 2006).

The question follows as to how to match these different instruments to the Danielson framework. Danielson and McGreal (2000) provide the chart in Table 2 to connect these various instruments to the framework (p. 53).

Table 2

*Danielson Framework Matched With Evaluation Instruments*

<table>
<thead>
<tr>
<th>Component of the Framework</th>
<th>Sample Sources of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1: Planning and Preparing</strong></td>
<td></td>
</tr>
<tr>
<td>1a: Demonstrating knowledge of content and pedagogy</td>
<td>Classroom observation; sample unit plan; sample lesson plan; interview; log</td>
</tr>
<tr>
<td>1b: Demonstrating knowledge of student</td>
<td>Interviews; sample lesson plan</td>
</tr>
<tr>
<td>1c: Selecting instructional goals</td>
<td>Sample unit plan/lesson plan; teacher artifact</td>
</tr>
<tr>
<td>1d: Demonstrating knowledge of resources</td>
<td>Sample unit plan/lesson plan</td>
</tr>
<tr>
<td>1e: Designing coherent instruction</td>
<td>Sample unit plan/lesson plan; teacher artifact</td>
</tr>
<tr>
<td>1f: Assessing student learning</td>
<td>Sample unit plan/lesson plan; teacher artifact</td>
</tr>
</tbody>
</table>
Table 2—Continued

<table>
<thead>
<tr>
<th>Component of the Framework</th>
<th>Sample Sources of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 2: The Classroom Environment</strong></td>
<td></td>
</tr>
<tr>
<td>2a: Creating an environment of respect and rapport</td>
<td>Classroom observation; student/parent survey</td>
</tr>
<tr>
<td>2b: Establishing a culture of learning</td>
<td>Classroom observation; teaching artifact, sample of student work; student/parent surveys</td>
</tr>
<tr>
<td>2c: Managing classroom procedures</td>
<td>Classroom observation; interview</td>
</tr>
<tr>
<td>2d: Managing student behavior</td>
<td>Classroom observation; interview; records of students sent to office</td>
</tr>
<tr>
<td>2e: Organizing physical space</td>
<td>Classroom observation</td>
</tr>
<tr>
<td><strong>Domain 3: Instruction</strong></td>
<td></td>
</tr>
<tr>
<td>3a: Communicating with students</td>
<td>Classroom observation</td>
</tr>
<tr>
<td>3b: Using questioning &amp; discussion techniques</td>
<td>Classroom observation</td>
</tr>
<tr>
<td>3c: Engaging students in learning</td>
<td>Classroom observation; teaching artifacts; samples of student work</td>
</tr>
<tr>
<td>3d: Using assessment in instruction</td>
<td>Classroom observation; samples of student work</td>
</tr>
<tr>
<td>3e: Demonstrating flexibility and Responsiveness</td>
<td>Classroom observation</td>
</tr>
</tbody>
</table>
Table 2—Continued

<table>
<thead>
<tr>
<th>Component of the Framework</th>
<th>Sample Sources of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 4: Professional Responsibilities</td>
<td></td>
</tr>
<tr>
<td>4a: Reflecting on teaching</td>
<td>Interview; reflection form</td>
</tr>
<tr>
<td>4b: Maintaining accurate records</td>
<td>Attendance records; field trip records</td>
</tr>
<tr>
<td>4c: Communicating with families</td>
<td>Phone logs; letters to parents; back-to-school night handouts; parent survey; parent phone calls</td>
</tr>
<tr>
<td>4d: Participating in a professional community</td>
<td>Logs of professional activities; copies of documents to which teacher has contributed</td>
</tr>
<tr>
<td>4e: Growing and developing professionally</td>
<td>Logs of professional goals and improved practice; copies of conference program attended</td>
</tr>
<tr>
<td>4f: Showing professionalism</td>
<td>Interview; feedback from colleagues</td>
</tr>
</tbody>
</table>

As mentioned earlier, a multidimensional teacher evaluation system examines teaching as a complex activity. Therefore, in order to evaluate its practice, multiple data sources and evaluators provide a more comprehensive understanding and perspective of what teachers do. While some of the evaluation criteria might be observable in the classroom, other measures such as communicating with families are invisible within the classroom; so other types of documentations and instruments are necessary. Consequently, in addition to principal evaluation of teachers through classroom observations, a multidimensional teacher evaluation system can also include the use of student achievement data, teacher portfolio, student evaluation of teachers, and peer review. These various components will allow for greater self-reflection and professional
growth by teachers who utilize portfolios, while also giving a more complete and holistic picture of teacher performance through the incorporation of student achievement data with principal, peer, and student evaluations.

The explanations of the evaluation instruments in the section below are by no means exhaustive. Nonetheless, major components of an evaluation system are listed with strengths, weaknesses, and major studies associated with each.

**Principal’s Role**

When examining teacher evaluation, the role of the principal must first be established. Current researchers mention that one of the primary responsibilities of a principal is to be an instructional leader (Camburn, Rowan, & Taylor, 2003; Mangin, 2007; Reitzug, West, & Angel, 2008). This role has become even more pertinent with the high-stakes accountability environment spawned by *No Child Left Behind* legislation (O’Donnell & White, 2005). As the instructional leader within a school building, principals regularly monitor instructional programs while also providing feedback to teachers in order to support teacher development (Reitzug, West, & Angel, 2008). However, the extent of this instructional leadership varies across elementary and secondary schools. Johnson and Holdaway (1990) discovered that at the elementary level, principals were more involved in the planning and supervision of instruction while secondary principals tended to delegate these responsibilities to others while influencing instruction indirectly.

The question follows as to the impact principals make within their schools as the instructional leader. Andrews and Soder (1987) conducted a study that showed student achievement gains were significantly greater in reading and math within schools that had
strong leaders versus schools with average or weak leadership. Later researchers have responded to this study by stating that principals do not have the same direct impact as teachers on student achievement; in fact, these researchers maintain that principals can indirectly influence student achievement by working with teachers (Hallinger & Heck, 1998; Heck, Larsen, Marcoulides, 1990; Siens & Ebmeier, 1996). Therefore, one of the main roles of principals as the instructional leader is to work with their staff to develop their teachers’ capacity as educators.

In working to develop teachers, principals also have the obligation as instructional leaders to intervene when marginal teaching exists (Kaye, 2004). During this intervention, principals take on varying roles such as mentoring and coaching and when necessary, they serve as counselors so that certain teachers are advised to move to another profession (Kaplan & Owings, 2001). At the same time, Peterson (2004) also emphasized the role that principals should take in recognizing and acknowledging effective teachers as well. In essence, principals as instructional leaders use the evaluation process as a means to improve teacher performance, whereby indirectly impacting student achievement (Toch & Rothman, 2008).

Another important responsibility of principals as instructional leaders is to create an environment of collaboration and shared leadership. For instance, an early study by Blase (1987) revealed that effective principals fostered positive relations between teachers, students, and parents. Abbott and McKnight (2010) more specifically discovered that collaborative schools positively influenced school leadership by strengthening principals’ instructional leadership. They came to this conclusion by pointing to studies that showed instruction steadily and continuously improved when
principals, as the instructional leader, provide stable school settings where collaborative and distributive leadership took place (McDougall, Saunders & Goldenberg, 2007; Saunders, Goldenberg & Gallimore, 2009). In other words, the distributive leadership style promoted positive outcomes such as better identification of student needs and appropriate classroom strategies, greater interaction across grade levels, and improved job satisfaction and teacher retention. Lastly, an extensive quantitative study funded by the Wallace Foundation of 4,165 teacher surveys in K-12 schools across the United States looked at the role of principals as instructional leaders (Wahlstrom & Louis, 2008). The researchers concluded with the following recommendation, “Our findings suggest that expanding the decision-making arenas in schools to include nonadministrators is an important step that leaders can take in long-term efforts to improve instruction” (Wahlstrom & Louis, 2008, p. 479). In other words, the study revealed that effective principals carry out their function of instructional leaders by practicing shared leadership. Consequently, principals as instructional leaders should take advantage of a multidimensional teacher evaluation system because this type of evaluation system correlates to the aforementioned call of instructional leaders to practice shared leadership. After all, a multidimensional teacher evaluation system incorporates multiple stakeholders in the evaluation process through the use of multiple measures and multiple evaluators.

Unfortunately, Brandt et al. (2007) discovered that most traditional evaluation systems involved the school principals and administrators as the sole conductor and evaluator. These evaluations are typically completed through classroom observations (McGreal, 1983). On a positive note, studies have shown that the use of classroom
observations by principals has the advantage of capturing information about teachers’ instructional practices (Mujis, 2006; Protheroe, 2002). After all, the classroom is where teachers teach. Peterson (2000) wrote that these classroom observations typically begin with a pre-conference with the teacher, followed by a formal observation, and finally a post-conference. Other times, principals may conduct an informal, unannounced classroom observation in order to obtain a quick and brief notion of what the teacher is doing in their classroom. This combination of announced and unannounced visits better ensures that observations capture a more comprehensive understanding of teacher practice.

Nevertheless, by utilizing this sole instrument by one evaluator only, studies have shown that teachers typically viewed these principal observations as “bureaucratic routine” (Holland, 2004) with little to no effect on teacher learning and improvement (Stronge, 2006). This is partially because inconsistent, brief observations by one person can create biased results (Medley & Coker, 1987; Wilkerson, Manatt, Rogers, & Maughan, 2000). Researchers suggest that when observations occur more frequently, its reliability improves (Denner, Miller, Newsome, & Birdsong, 2002). Therefore, classroom observations should take place regularly rather than one or two times a year, as is the case with traditional evaluation systems.

Furthermore, by including multiple evaluators in the observation process, credibility improves. As Toch & Rothman (2008) wrote, “another way to counter the limited, subjective nature of many conventional evaluations is to subject teachers to multiple evaluations by multiple evaluators” (p. 7). As an example, Lower’s (1987) study of 63 school districts in Ohio revealed that when multiple evaluators were utilized
in the evaluation system, teachers and even the principals had more confidence in the evaluation process and perceived that the evaluative information was more useful in improving teacher performance.

There are other reasons for teachers’ dissatisfaction with traditional evaluations conducted by principals. Brandt’s et al. (2007) study of evaluation policies in several Midwestern districts found that most of the principal evaluations were summative in nature, with little formative assessment taking place to develop teachers. Secondly, several researchers have discovered that principals are too busy to conduct through and detailed evaluations of their teachers (Copland, 2001; Danielson, 2007; Donaldson, 2010). As an example, Zimmerman and Deckert-Pelton (2003) conducted a study across five Florida county school districts to examine teachers’ perceptions of their principal as evaluators. Though teachers consistently expressed their desire that the evaluation process should be a bidirectional process of constructive principal feedback, many expressed that it did not take place. Several teachers articulated a major reason for this deficiency by responding, “‘many times the principal is not able to spend an adequate or equivalent amount of time in the classroom evaluating teachers’” (p. 34). This is because administrative tasks typically take away from the principal’s time for classroom observations and conferences with teachers. After all, evaluating teachers is one of many responsibilities for principals, who handle multiple tasks such as scheduling, budgeting, reporting, communicating with parents and community, and handling unforeseen circumstances that are inevitable within a school building (Fink & Resnick, 2001).

Besides principals lacking time and focus on developing their staff, many teachers also believed that their principals lacked the necessary expertise to conduct observations
(Danielson & McGreal, 2000). Brandt’s et al. (2007) study discovered one reason for this perception; the study found that only 8% of Midwestern districts mentioned evaluator training as a component of their teacher evaluation system. Another study by Jacob & Lefgren (2008) examined the accuracy of principal evaluations by comparing it to actual student achievement scores. The researchers found that principals were able to identify effective teachers in the top or bottom band but were less successful at differentiating between teachers in the middle band. A final study by Wilkerson, Manatt, Rogers, and Maughan (2000) discovered that student ratings of teachers were better predictors of student achievement than principal ratings. In essence, there are noticeable issues related to training and expertise of principals to conduct teacher evaluations.

All of these deficiencies in principals as evaluators further attest to the need for the administrators, as the instructional leader, to take advantage of a multidimensional teacher evaluation system. This system includes classroom observations by the principal while also incorporating other evaluation instruments as well. Before looking at the benefits of including other instruments, it is first necessary to tackle the concerns regarding principals as evaluators. To begin with, the belief that principals lack expertise in conducting observations calls for the need in establishing thorough evaluator training. Muijs’ (2006) study specifically recommended this crucial element of rater training when using the evaluation instrument of classroom observations. The training should familiarize principals with the evaluation instrument, taking them meticulously through all the steps. Plus, continual practice must take place in order to reach a properly high level of reliability.
Even if extensive training is given to principals, Jacob and Lefgen (2006) advised against exclusively using principal evaluations because their study showed that administrators still had an "inability to distinguish between a broad middle range of teacher quality," which “suggests caution in relying on principals for fine-grained performance determinations" (p. 5). In these cases, Yon, Burnap, and Kohut (2002) suggested utilizing peer evaluators and content experts who can better assess and differentiate teacher practice since principals may not have the specialized knowledge to make detailed observations. Furthermore, these peer evaluators are typically given exclusive time to focus on providing teacher feedback and therefore can also help fill the gap in fulfilling the formative function lacking with administrators, who may be deficient in expertise and time. In fact, as Van Zant, Raszka, & Kutzner (2001) explained, these peer evaluators are able to serve as effective mentors since they are themselves professional teachers with many years of experience.

In summary, an important role of principals as the instructional leader is to utilize more data sources, such as peer review, in the evaluation process in order to fulfill both formative and summative purposes while at the same time involving multiple stakeholders in the evaluation process. By doing so, principals also fulfill their responsibility as instructional leaders in creating an environment of collaboration and shared leadership (Abbott & McKnight, 2010; Wahlstrom & Louis, 2008). Finally, by incorporating other data sources and evaluators, the principal assures greater reliability and reduces biases in their documentation of teacher performance (Stronge, 2006). In essence, one of the major roles of the principal is to take the lead as the instructional head in developing and implementing a multidimensional teacher evaluation system. After all,
McEwan (2003) wrote, “Effective instructional leaders take responsibility for making sure that trustworthy research and proven practices are talked about frequently and demonstrated ably in their schools” (p. 36). Goldstein’s (2004) study also revealed that teachers felt that a properly trained administrator should be the leader in implementing the evaluation process by looking at multiple sources of data. Such data incorporates principal evaluation of teachers with other instruments such as student achievement assessments, evaluations by other assessors such as peer teachers and student, and teacher self-reflection through portfolios.

**Student Achievement Data**

As mentioned earlier, the task of evaluating teachers involves different components due to the fact that teaching is a complex and multifaceted task (Peterson, Stevens, & Ponzio, 1998). Instruments such as the *Framework for Teaching* and teacher observations look primary at what teachers do; these instruments are focused on input data. Danielson (2008) pointed to another key component when evaluating teachers: “teachers should be evaluated on the basis of the results they achieve with students—on their outputs” (p. 55). It has already been established through strong empirical studies that teachers are the most important school-based factor of student achievement (Borman & Kimball, 2005; Marzano, 2003; Wright, Horn, & Sanders, 1997). Teaching is, in the end, about helping students learn. Unfortunately, teacher’s effectiveness of carrying out the task of educating children is not often measured in relation to their effectiveness in raising students’ tested achievement (Vigdor, 2008). As examples, studies listed earlier point to the ineffectiveness of traditional teacher evaluation systems that fail to differentiate teacher effectiveness; as Weisberg et al. (2009) pointed out, “over the last
three years, only 10 percent of failing schools issued at least one unsatisfactory rating to a tenured teacher” (p. 12). Consequently, traditional evaluation systems fail to recognize superior teaching performance and develop weaker teachers.

Policymakers and researchers are therefore insisting that school districts should include data that shows teachers’ ability to demonstrate student academic growth as a part of the overall teacher evaluations system. At the federal level, President Obama’s Race to the Top grant competition placed a top priority in using student achievement growth to reward effective teachers. Because of this directive, many states revised their laws to include test scores within their evaluation systems with the hope of receiving a portion of the $4 billion in Race to the Top funding (Frist, 2010). In Michigan, lawmakers passed key education reforms in December 2009, which included the following section: “connect student performance as one component in teacher and administrator evaluations” (Michigan, 2009).

Furthermore, private donors involved in teacher evaluation reforms, such as the Bill and Melinda Gates Foundation and the Carnegie Foundation, have also placed top priority in incorporated student achievement data in their evaluation systems. An essential premise of the Gates Foundation’s project to develop high-quality teacher evaluation systems known as the Measures of Effective Teaching project, with funding of $45 million, is: “First, whenever feasible, a teacher’s evaluation should include his or her students’ achievement gains” (MET, 2009). Additionally, the Carnegie Foundation is the major funder for the Operation Public Education program which focuses on promoting teacher performance by primarily focusing on student achievement scores (Center, 2004).
The question then follows as to how to measure student achievement. One method is through standardized tests. Sanders and Horn (1995) provided many positive qualities of using standardized tests. Besides measuring basic skills, standardized tests can even measure higher level thinking skills if it is well designed. Plus, these tests have the advantage of being written by specialists with standardization of administration and scoring procedures, which allows for comparisons with national norms and students. It can also cover a number of subjects so that a good overview of the students’ knowledge of the curriculum is provided. It therefore seems to make sense that teacher effectiveness should be measured in part by standardized test scores of students.

Toch and Rothman (2008) revealed a major problem with this line of thinking. Teachers receive different types of students. Some teachers work with students with privileged background while other teachers are dealt students who are less prepared. Even if the teacher of well-prepared students underperforms, they may still score high on the standardized test while less fortunate students may score lower despite the teacher’s hard work to get them back to grade level. As one example, the Center for Greater Philadelphia pointed out that SAT scores rise with every $10,000 of family income; this should not be surprising since variables of higher income families such as good jobs, years of schooling, positive attitudes about education, the capacity to expose one’s children to books and travel, etc. correlate to higher test scores (Center, 2004). “So handing teachers high ratings if their students have high test scores would result in many misjudgments of teacher’s true abilities” (Toch & Rothman, 2008, p. 4). Looking at this reality differently, achievement scores between students at the end of the year are highly correlated with differences between those students at the beginning of the year. These
differences affecting test scores can include factors such as parent’s educational level or socio-economic status, which are beyond the control of teachers.

Therefore, controlling for these various outside factors would lead to a focus on student growth during the academic year. In following this line of thinking, researchers point to the use of value-added assessment to accomplish this purpose (Ballou, Sanders, & Wright, 2004; Hershberg, 2005; Sanders, 2000).

Harris (2010) defined value-added assessment as a method for measuring growth of a student during a school year. Instead of measuring levels of absolute achievement, the student’s academic growth is measured by subtracting initial student achievement test score from the end-of-year score (Hershberg, Simon, & Lea-Kruger, 2004). This growth measure allows one to make a better prediction of future student achievement. After all, a better way to predict how a student will do next year is by looking at how they did last year. Harris (2010) defined value-added as a “statistical technique that makes predictions based on information that includes but goes beyond prior student achievement” (p. 67). In essence, these predictions serve as the basis for comparison to what actually happens. This means, for instance, that high value-added scores come from students who receive scores higher than predicted. Because this approach focuses on individual students traced over time, “each student serves as his or her own ‘baseline’ or control, which removes virtually all of the influence of the unvarying characteristics of the student, such as race or socioeconomic factors” (Hershberg, 2005). In summary, value-added assessment takes away outside influences of students that are beyond teacher control in order to focus on the impact of student learning during a school year.
One of the concerns with using value-added assessment is its reliability. For instance, McCaffrey, Sass, Lockwood, and Mihaly (2009) referred to measurement errors that may appear in scores from year to year due to different factors that lead to measured scores differing from real scores. These researcher go on to explain that these sampling errors tend to average out across a teacher’s total number of students; so the sampling error tends to fall with the number of students per teacher per year. Other factors for variation of scores over time come from “nonpersistent changes in performance” (McCaffrey, Sass, Lockwood, & Mihaly, 2009, p. 578). These possible errors refer to many variables such as the test itself, dealing with the selection of certain questions and familiarity with tests (Hanushek & Rivkin, 2010). Other variables include effects of different groups of students in teacher-classrooms from one cohort to another and changes in school administration and instructional programs (Bock, Wolfe, & Fisher, 1996). To combat these various issues, Ballou, Sanders, and Wright (2004) produced a study that showed more accurate estimation of teacher effectiveness that reduces these types of errors by including multiple years of student achievement data. An example of this method is the Tennessee Value-Added Assessment System, which attempts to include three to five years of students’ prior achievement data when estimating their teachers’ value-added measurement (Tennessee, 2004). In addition, McCaffrey, Sass, Lockwood, and Mihaly (2009) recommended averaging teacher effectiveness estimates across two to three years, helping to further reduce these measurement errors. The Tennessee Value-Added Assessment System also includes this approach by evaluating gains for individual teachers in terms of average gains over the three most recent years.
These types of strategies essentially provide greater reliability to a value-added assessment system.

A recent study by Chetty, Friedman, and Rockoff (2011) looked at the relevance of using teacher value-added measurements. One complaint about its use is the possibility of bias where teachers receive higher or lower value-added measurements based upon sorting of their students (Rothstein, 2010). The researchers used data covering more than 2.5 million students from a large urban school district in the U.S. from grades 3-8. Scores from math and reading tests were examined for these students from 1989-2009. To evaluate if value-added estimates are biased, parent characteristics from tax data were also used across the same time period. Results showed that parent characteristics were uncorrelated with teacher value-added measurements. Furthermore, the researchers then looked at whether high value-added teachers improve their student’s outcomes into adulthood or were these teachers simply better at teaching to the test. Results showed that teacher value-added measurements substantially impacted students on a board range of outcomes. For instance, “A 1 SD improvement in teacher value-added in a single grade raises the probability of college attendance at age 20 by 0.5 percentage points, relative to a sample mean of 36%” (Chetty, Friedman, & Rockoff, 2011, p. 3). Additionally, students of higher value-added teachers also showed steeper earnings trajectory. Essentially, the study showed that value-added measurements provide evidence of teacher’s long-term impact on their students.

On the other hand, there are major concerns in using value-added assessment on a practical level. First, researchers such as Peterson (2004) have pointed to the lack of standardized tests for all teachers, which prevent calculation of value-added estimates.
To help answer this problem, Sawchuk (2011) referred to the use of student learning objectives as the most heavily studied alternative for measuring teacher impact on learning. For instance, Denver Public Schools allows teachers without standardized test scores to set two student growth goals for the year that are approved by their supervisor (Denver, 2010). These goals should be based upon performance objectives that are measured through other types of assessments such as benchmark or curriculum tests. Alternatively, the District of Columbia Public Schools (2009) allows teachers to use appropriate commercial assessments at beginning and end of year for non-tested grades and subjects. These commercial tests usually have the advantage of already being tested for issues of fairness and appropriateness. And in the absence of finding substitute assessments, the District of Columbia Public Schools provides a substitute evaluation system that places a greater proportion of 75% of evaluation scores based upon administrator and master teacher observations.

Another practical concern of using value-added assessment is the students. In mostly urban districts, a high rate of student mobility is a continual problem that teachers have faced for many years (Ingersoll, Scamman & Eckerling, 1989). So how can teachers be held responsibility for student that they had for only a portion of the year? The Tennessee Value-Added Assessment System provides a possible answer with their policy that a student who is enrolled in a teacher’s class for at least 75 days in a one-semester course or 150 days per year will have 100% of their test performance matched to that teacher; otherwise, 50% is attributed to the teachers if the student is enrolled 75-149 days, while a student enrolled fewer than 75 days of the school year will not have their test results matched with the teacher (Tennessee, 2004). Essentially, these examples
show how school districts are able to find approaches to help resolve practical issues when incorporating value-added assessments to their teacher evaluation system.

On the other hand, value-added assessments have a number of limitations that show that it should not be used as the sole basis for making evaluation decisions. First of all, one of the risks in exclusively looking at test scores is the likelihood that teachers will engage in test-taking instructions instead of teaching to the entire curriculum (Sloane & Kelly, 2003). Another term for this methodology is teaching to the test. A possible result with this type of approach is stifling creativity of students who are taught how to answer multiple choice questions through continuous drills. Longo (2010) specifically found examples of teacher instruction in science conducted through test-taking drills where answers are provided to students instead of the researcher’s preferred technique of inquiry-based learning. Diamond’s (2007) study through observations and interviews with teacher showed that high-stake testing led to didactic instructions dominating instead of interactive learning, especially in low-income schools. These teacher-led instruction emphasized lectures, where student involvement occurred primarily through reciting answers and seat work. Finally, Booher-Jennings’ (2005) study discovered that schools that are extensively focused on testing led to teachers engaged in dishonest tactics such as referring lower achieving students for special education to increase their results. The researcher noted that even though certain schools did not tie any consequences to teachers based on test scores, teachers nonetheless engaged in these practices because gains in test scores was understood almost exclusively as a reflection of teacher effectiveness. These examples show the importance of not placing reliance on student achievement scores alone.
It should be noted that options besides standardized testing alone are available when assessing student learning. For instance, the National Center for Fair and Open Testing have called for greater use of local and classroom evidence of learning based on students’ work. “By focusing on the classroom, we can assess important learning standardized tests cannot, such as research projects, oral presentations, essays, problem solving and using computers in real situation” (FairTest, 2010, p. 1). These types of assessments allow for the evaluation of students’ higher level and deeper thinking skills, which standardized testing may not provide. But these types of assessments also require certain additional steps to ensure quality in scoring, such as extensive training for assessors and independent graders to verify initial scoring.

Furthermore, teacher effectiveness cannot be measured solely by academic tests. Darling-Hammond (2004) acknowledged that the movement towards increased accountability in education can lead to an overreliance on testing; she wrote, “indicators such as test scores are information for the accountability system; they are not the system itself.” (p. 1081). Furthermore, Campbell, Kyriakides, Muijs, and Robinson (2004) noted, “Teacher effectiveness is the impact that classroom factors, such as teaching methods, teacher expectations, classroom organisation, and the use of classroom resources have on student performance” (p. 2). These other factors can only be truly captured when value-added assessment is combined with other evaluation instruments such as teacher observations. Additionally, evaluation instruments beside student achievement data will be necessary for teachers who meet any one of three conditions: those who teach subjects that are not tested annually, those who do not have multiple years of achievement data available, or those who have highly mobile students.
And finally, though student achievement scores help to fulfill the summative purpose of an evaluation system, when it is used solely, it is unlikely to meet the formative function. For instance, Rivkin, Hanushek, and Kain’s (2005) study pointed out that student achievement scores revealed that teachers vary in their contribution to their students’ achievement score gains, but the scores itself cannot explain what caused the variation. In order to discover reasons for these variations, other instruments are needed such as observations, peer review, student feedback, and teacher portfolios. Peterson (2006) further pointed out that these other data sources can also provide additional evidence to student achievement gains by showing “that the content taught is of high quality” through peer review of materials, for instance, and “that students have positive attitudes” through student evaluation of teachers (p. 218). In summary, all these factors speak to the importance of including other instruments in addition to value-added assessments to a multidimensional teacher evaluation system.

**Teacher Portfolio**

According to Danielson (2001), one of the common complaints about traditional teacher evaluation system is that it is “an activity that is done to teachers” (p. 13), where educators play a passive role. Another deficiency mentioned earlier by Danielson and McGreal (2000) is its failure to have a positive impact on professional growth and leadership practices. Both of these shortcomings are mitigated through the incorporation of a teacher portfolio.

A portfolio is defined by Beck and Weiland (2001) as “a systematic collection of materials or artifacts that demonstrate the capabilities of the person who compiled the portfolio” (p. 60). A portfolio-based appraisal system includes a self-assessment
component that allows teachers to continually analyze their practice as it relates to their annual goals. This analysis takes place with the teacher and administrator first reviewing together the teacher’s annual goals. Then, during the summative stage, the teacher presents their portfolio to their evaluator, displaying their accomplishments throughout the year. Clarification and feedback by the evaluator takes place during this time, and the final appraisal rating is based on their collaboratively-developed rubrics (Danielson, 2001). The process does not end here because refocusing and planning occurs soon thereafter, calling for continued professional growth aimed at improvement for the coming year.

At the heart of a portfolio-based appraisal system is a collection of artifacts and reflective entries that represent teacher growth, which are directly related to their annual goals. Artifacts include different types of information such as “teacher-reflective journal entries, documentation of student activities, videotaped teaching lessons, lesson plans, surveys, newsletters, changes in curriculum, peer observation, and conference materials” (Beck & Weiland, 2001, p. 62). As teachers gather and select artifacts, they become active researchers who continually analyze and reflect on their own practices (Attinello, Lare, & Source, 2006; Beck & Weiland, 2001; Danielson, 2001; Hackmann & Alsbury, 2005).

Studies have also shown other benefits of using a teacher portfolio. Besides the teacher becoming an active self-reflecting participant, a portfolio-based appraisal system has also shown positive impact on the two areas of teachers’ professional growth and leadership practices (Shulman, 1988; Wolf, 1999). The possibility to provide feedback on these two areas of teacher practice is particularly important because they are not
readily observable during classroom visits. As an example, a multiyear mixed design study within a county school district in Virginia that used portfolios in their teacher review for three years showed that it enhances the accountability and professional development purposes of teacher evaluation such that teachers in the focus group discussions commented, “[The portfolio] helps me know where I need and want to go. It keeps us from becoming stagnant” (Tucker, Stronge, Gareis, & Beers, 2003, p. 591). Furthermore, data from surveys and focus groups from the same study indicated that teachers and administrators also viewed portfolios as fair and accurate measures of performance; in fact, the highest ratings by both administrators and teachers were in these two areas of fairness and accuracy. As an extension of studies that looked at portfolio's level of accuracy and reliability, Peterson (1988) conducted a study looking at the reliability of using teacher dossiers, which is a more compact form of teacher portfolios. He found that it can be reliable for making summative evaluation decisions since the 26 evaluators in the study showed a 90.4% agreement in evaluation judgments across the various teacher dossiers. Peterson's (1990) subsequent study also showed that dossiers were reliable in making simulated rankings of teacher performance.

In summary, these examples show that incorporating a portfolio-based appraisal system makes the evaluation process more relevant and worthwhile by allowing teachers to become active participants in the process. It also addresses the situational and complex nature of teaching by personalizing the portfolio with individual goals for professional growth. Lastly, portfolios are flexibility enough to be used with diverse student population from all grade levels and subjects (Beck & Weiland, 2001).
On the other hand, evaluating teachers based on their portfolios has not shown conclusive correlation to student achievement gains. For instance, the National Board for Professional Teaching Standards (NBPTS) is a program that uses portfolio assessment to evaluate teaching. Studies connecting NBPTS certification with student achievement have shown assorted results (Cavalluzzo, 2004; Goldhaber & Anthony, 2007; Pool, Ellett, Schiavone, & Carey-Lewis, 2001). Furthermore, no conclusive findings exist on the reliability of portfolio assessments as a part of an objective teacher evaluation system (Attinello, Lare, & Waters, 2006). Secondly, an additional problem is that though portfolios provide rich data about the teacher, the data are known to be difficult to score (Wolf & Dietz, 1998). This is because a portfolio’s advantage of being open-ended can lead to a plethora and variety of artifacts that become too difficult to make sense of. In order to help resolve this concern, Wolf, Lichtenstein, and Stevenson (1997) suggest including specific guidelines and procedures beforehand to ensure manageability and fairness; examples include the “identification of sound content and performance standards for teachers,” which will serve to guide teachers and reviewers in the construction and evaluation of portfolios, and also “specification of the requirement for construction of a portfolio” so that the portfolio is focused on a few key areas of teaching (p. 172). Van Wagenen and Hibbard (1998) also discovered that portfolios have greater value when they are more specific and limited in scope, which makes it more tailored to meet the development needs of each individual teacher.

Another concern with portfolios relate to its reliability. This is because teachers are the ones who choose which of these artifacts to include, thus introducing self-report bias and possible misrepresentation (Wolf & Dietz, 1998). Questions can therefore be
raised as to whether the portfolio accurately reflects what occurs in classrooms. As one teacher said within Attinello, Lare, and Waters’ (2006) study, “I could put together a really nice portfolio and not be a very good teacher. Conversely, a great teacher might not create a good portfolio” (p. 141).

Finally, a couple of practical concerns exist when using teacher portfolios. First of all, evaluators oftentimes will need to know the content area in order to assess grade and subject-level artifacts. Incorporating teacher peer review helps in resolving this problem. Another practical concern expressed by many teachers and administrators in studies related to portfolios is the burden of time needed for its development and implementation (Attinello, Lare, & Waters, 2006; St. Maurice & Shaw, 2004; Tucker, Stronge, Gareis, & Beers, 2003). This concern therefore entails the need for build-in time for teacher and administrators when implementing portfolios as a part of the school’s teacher evaluation system. Additionally, some researchers have answered this concern of time by pointing to the benefits of using an electronic portfolio system. For instance, a qualitative study by Jun, Anthony, Achrazoglou, and Coghill-Behrends (2007) indicated that 90% of administrators who used electronic portfolios responded that they would encourage their teachers to build and maintain ePortfolios because it offers them the advantage of having constant access to their teachers' samples while also giving teachers prompt feedback. Other teacher benefits listed in the same study included an efficient organizational scheme, ample storage space, portability, interactive presentation capability, and multiple linkage opportunities. Another quantitative study by Pecheone, Pigg, Chung, and Souviney (2005) found similar administrator and teacher feedbacks related to time savings since electronic portfolios provide real-time feedback from any
computer via the web. On the other hand, these studies and others cautioned for the need in adequate training for both evaluators and evaluatee and also the learning curve involved when incorporating electronic portfolios (Bartlett, 2002; Heath, 2005).

Due to these concerns of using portfolios, especially regarding its accuracy and reliability, studies have recommended using it inclusively but not exclusively within an evaluation system (Attinello, Lare, & Waters, 2006; Tucker, Stronge, Gareis, & Beers, 2003). In fact, portfolios can serve as a complement to data collected through other instruments such as classroom observation and student achievement since it provides a more complete picture of the teacher. As an example, a previously mentioned study by Tucker, Stronge, Gareis, and Beers (2003) of a county district in Virginia found that teacher portfolios helped to more precisely differentiate quality of teacher performance. Before its incorporation, the district rated 96% of their instructional staff as above average for a number of years. With the introduction of portfolios, greater differentiation in evaluation ratings was evident with 43% of teacher rated as above average and 55% rated as meeting expectations. In other words, by incorporating the additional instrument of teacher portfolios, it helped evaluators differentiate better between the competent teachers and the extraordinary ones. So the evidence presented by the teacher’s portfolio will serve as additional data to actual classroom practice and student outcomes. However, Peterson (2000) also reminded his readers that though portfolios can provide greater perspectives of teacher practice, it nonetheless falls short in revealing other strengths of teachers such as personal interactions that are difficult to show through tangible artifacts. Hence, Danielson (1996) concluded, “A portfolio is not a substitute for
classroom observation and professional discussion . . . Rather, it is an extension and enhancement of that discussion to cover all aspects of teaching” (p. 39).

**Student Evaluation of Teachers**

Besides a portfolio-based appraisal component, a multidimensional teacher evaluation system should also look at other sources of information in order to obtain a more complete and holistic view of the teacher. As Epstein (1985) wrote, “because there is no single set of skills that perfectly define effective teaching, measure of many aspects of teaching by multiple judges are likely to yield the fairest and most comprehensive evaluation of teachers” (p. 8). Therefore, another valuable source of information worth considering for a multidimensional teacher evaluation would be the students since they are the ones most present when teachers are teaching. As the saying goes about teacher evaluation, “‘ask the kids, they know who the good teachers are’” (Peterson, 2000, p. 103). It should also be pointed out that opinions from students are already being used by administrators for informal teacher evaluations on a day to day basis. By incorporating formal student evaluations, it makes these informal ongoing evaluations more accurate through research-supported surveys completed by a larger student population. Otherwise, the alternative “means that this part of teacher quality documentation will be abandoned to such things as unreliable observations, happenstance events, gossip, hearsay, malicious individuals, teacher uncertainty, and unsystematic Internet teacher abuse sites posing as acceptable rating services” (Peterson & Peterson, 2006, p. 49).

While student evaluations of teachers are traditionally done at the college and post-graduate level, there are also recorded instances where it is used as early as the elementary levels (Mathews, 2000; Peterson, Driscoll, & Stevens, 1990). This is because
studies have shown links between positive student evaluations of teachers and effective teaching as measured through gains in student achievement (Cohen, 1981; Dickinson, 1990; Johnson, 2012; Kyriakides, 2005; Tevin & McCrosky, 1997; Yonghong & Chongde, 2006). In fact, teachers seem to agree that student evaluations are a useful and appropriate mechanism of feedback about their performance in the classroom because it provides them with information about their instructional skills, classroom demeanor, teaching strategies, as well as student learning outcomes (Parayitam, Desai, & Phelps, 2007; Peterson, Wahlquist, Bone, 2000; Simpson & Siguaw, 2000). After all, as Follman (1992) pointed out, students are the most direct clients of teachers and therefore have a broader and deeper experience with their instructors than other potential evaluators such as administrators, peers or parents. An interesting study examined this belief by looking into the relationship that exists between performance of K-12 students on criterion-referenced reading, language arts, and mathematics tests with teacher performance measures made by three participants: principals, students, and teachers who rated themselves (Wilkerson, Manatt, Rogers & Maughan, 2000). Questionnaires were given for four rater groups: lower elementary (K-2), upper elementary (3-5), middle school (6-8), and high school (9-12). Results showed that the highest positive correlation existed among measures of student achievement and feedback from students with 0.67 in math, 0.75 in reading, and 0.70 in language arts. Furthermore, the same study found that student ratings were the best predictor of student achievement on the criterion-referenced test, with significant results seen in all three subject areas using stepwise multiple regression analysis.
One of the complaints often heard about student evaluations is the potential for this tool to be manipulated by teachers who give higher grades or easier coursework in order to receive better evaluations. One particular study conducted by McPherson (2006) looked at data that covered eight and a half academic years across 607 economics classes. His conclusion after conducting regression analysis was that instructors can buy higher student evaluation scores by awarding higher grades. But one can also make the case that higher grades are correlated with higher student ratings of teachers since these students are satisfied with what they learned from the instructor. Additionally, this same study that was meant to invalidate student ratings also concluded from their analysis that the level of experience of the instructor and the class size were significant determinants of student evaluation scores; this conclusion actually provides further validation for the use of student evaluations given that lower class size and greater instructor experience have been associated with more effective teaching (Arubayi, 1987; Feldman, 1984). Another study by Centra (2003) looked into this same debate on the influence of grades and coursework on student evaluations. Centra’s study examined over 50,000 courses across eight subject areas. After controlling for factors such as class size, teaching method, and student perceived learning outcomes, expected student grades generally were shown not to affect student evaluations. “In fact, contrary to what some faculty think, courses in natural sciences with expected grades of A were rated lower, not higher” (Centra, 2003, p. 495). In addition, courses were rated lower by students who rated the class as either difficult or too elementary. Hence, the highest evaluation scores were given by students who thought the class was just right. The study concluded by stating, “What these findings indicate is that teachers will receive better evaluations when their courses are
manageable for students” (Centra, 2003, p. 515). In other words, higher grades or easier coursework does not necessarily translate into better student evaluations scores.

Another argument against student evaluations is the concern for student biases towards their teacher based upon factors ranging from gender to teacher likeability. To look into this issue, Ebmeier, Jenkins, and Crawford (1991) conducted a comparison of high school students’ ratings of meritorious and non-meritorious teachers with the ratings from qualified evaluators. The study found that students were able to discriminate between the two groups as well as the expert practitioners. Furthermore, Peterson, Wahlquist, and Bone’s (2000) study involving 9,765 students concluded that students are able to “distinguish between merely liking a teacher and recognizing one who enables their learning” (p. 148). Moreover, Emery, Kramer, and Tian (2003) pointed out that even if a few students have certain teacher biases or fill out the evaluation without the purest intentions, it is important to note that the aggregate scores should considered when evaluating feedback from students; in other words, outliers are diminished by the aggregate. Peterson and Stevens (1988) similarly wrote that the larger number of student population completing a student evaluation provides high reliability for measuring teacher performance. In fact, they showed evidence of class data reliability in the 0.80 to 0.90 range (Peterson & Stevens, 1988).

As with any data source, certain safeguards should be set in place to control the use of student surveys. In fact, it can be argued that unethical and inappropriate behaviors should not be used as the reason to invalidate the accuracy of the process. Theall (2010) wrote:
Calling the basic process invalid because it has been manipulated or because it is poorly done, or because data are misused, or calling the providers of data biased is a mistake because the preponderance of evidence has shown that when properly done and properly used, [students’] ratings data can be accurate, reliable, and useful (p. 45).

On the other hand, steps should be taken to protect the integrity of the process as well. First of all, individual teachers should not develop and administer their own surveys. Standardization of research-proven and field-tested survey forms produced by committees with some level of customization can aid fairness and useful interpretation (Peterson, 2000). In regards to the makeup of the appraisal system of student evaluation, it must first be multidimensional so that it provides meaningful data about areas such as the content of the course, delivery style of the teacher, pacing, amount and relevance of the workload, and learning outcomes within a particular class (Abrami & D’Apollonia, 1999). And equally important is the student suggestions on how to improve any of these components so that teacher effectiveness may be enhanced in the future. For primary grades, student evaluations are more commonly distributed at the end of a unit of instruction or lesson with forms that are typically short and age-appropriate (Mathews, 2000). High school students complete a more comprehensive evaluation form that Baldwin (2004) stated should include “a series of multiple choice questions followed by the invitation for students to write a few paragraphs or a short essay” about their teacher (p. 102).

Regardless of the grade level in which student evaluation of teachers are given, it can nonetheless be used to meet the formative and summative functions of an evaluation
system. First of all, it is important that data from student evaluations is shared with teachers in a timely fashion. Since the expectation from teachers is that they reflect upon student feedback, the evaluations should be shared as soon as possible in order for the teacher to have the opportunity to modify their strategies and practices (Danielson, 2003). Hence, student evaluations of teachers can be used for self-reflective purposes, giving teachers opportunity to consider ways and means to possibly modify their instruction and practice. This meets the formative function of an evaluation system.

Secondly, in regards to summative assessments, researchers caution the use of student feedback for this purpose. First of all, Peterson and Stevens (1988) suggested that student data for several years are first necessary in order to establish patterns of teacher performance. Peterson (2000) wrote elsewhere, “Years of experience for both students and teachers are required for teacher evaluation systems to acquire their full functioning and utility” (p. 120). Only thereafter should student ratings be considered as an instrument in fulfilling the summative portion of a teacher evaluation system. And even then, Danielson and McGreal (2000) wrote that student evaluations should serve as a supplement to other indicators of teacher performance.

There are some final benefits of using student evaluations within a multidimensional teacher evaluation system. It allows teachers to see their classroom through the eyes of their students. Plus, student evaluations help in cultivating a feeling of community where students’ thoughts and feedback are valued (Gabriel, 2005). It must be mentioned here that these feedbacks should take place throughout the school year and not only at the end of course so that teachers can make mid-course corrections. This type of appraisal system allows teachers to reflect on their practice continually. Finally,
Worrell and Kutherbach (2001) wrote that student evaluations have the added advantage of being cost and time-efficient with minimal obstruction and training needed for those involved in the process.

On the other hand, researchers caution that student evaluations should not be used primarily to measure teacher effectiveness since students are not able to comprehensively evaluate teachers on content knowledge, curriculum, and other realms of effective teaching (Emery, Kramer, & Tian, 2003; Follman, 1992; Worrell & Kuterbach, 2001). Furthermore, Simpson and Siguaw (2000) argued for using other instruments to evaluate teachers so that possible manipulations by teachers and biases by students are further mitigated by multiple perspectives, whereby bringing greater reliability and validity to the process. With these various safeguards set in place, student evaluations of teachers can provide a more comprehensive and multidimensional evaluation of teacher performance.

**Peer Review**

In extension of the idea presented in the previous section that a multidimensional teacher evaluation should take into consideration perspectives from different individuals, feedback from peer teachers can consequently be an important component within an evaluation system. After all, the expertise of teachers within a school should be utilized as much as possible. “No one else in the educational system practices the skills of combining knowledge of students, application of curriculum, and supporting student achievement quite like classroom teachers” (Peterson, Kelly, Caskey, 2002, p. 318).

Peer review is also labeled as peer observation or peer coaching because the process is meant to improve teacher quality through the support of colleagues (Olebe,
Jackson, Danielson, 1999). This type of support meets the formative function of an evaluation system and it extends to all teachers in different stages in their career. For instance, the Poway Professional Assistance Program provides support for beginning teachers through full-time teacher consultants who regularly meet with them, while second-year and transitional teachers are given on-site support from veteran teachers, and finally experienced, tenured teachers set their own professional goals in conjunction with completing peer coaching and classroom visitations within their school building (Van Zant, Raszka, & Kutzner, 2001). These types of resources and opportunities allow teachers to grow as professionals as they self-reflect on their practices. Another well-known peer review program is the Toledo Plan that goes a step further by incorporates peer assistance for veteran teachers who are identified as struggling (Hertling, 1999). And for beginning teachers, the program incorporates an internship before the start of the school year, where they are matched with an expert consulting teacher for the purpose of professional development and evaluation. In essence, both of these programs utilize the expertise of peer teachers for those who need who need it. In fact, Van Zant, Raszka, and Kutzner (2001) compared the programs to a residency program in medical school where teachers receive hands-on and real-time training from peers; the researchers concluded, “Training is completed in the best place for new teachers to learn – their own classrooms” (p. 22).

It is again worth noting that all teachers involved throughout the peer review process are given opportunities for self-growth, regardless of whether the teacher is one being evaluated or the one doing the evaluation. This concept stands in contrast to the experience of traditional beginning teachers who are oftentimes placed in a classroom
after as little as one semester of student teaching, with little or no support (Veenman, 1984). By incorporating peer reviews, expert teachers with many years of experience now provide mentoring for these teachers. “The consultant provides the new teacher with thorough and frequent classroom observations, reflective comments, curriculum support, model lessons, coaching, classroom arrangements, and sometimes just a sympathetic ear” (Van Zant, Raszka, & Kutzner, 2001, p. 22). For veteran teachers, self-growth occurs as they take part in professional development trainings in order to become mentors for beginning teachers. Throughout this process, they develop skills in examining teaching techniques and assessing these techniques through classroom observations. Veteran teachers who took part in the California Formative Assessment and Support System for Teachers “often comment that their own learning about teaching is as great or greater than that of the beginning teachers” because of the invaluable trainings they received before becoming mentors (Olebe, Jackson, & Danielson, 1999, p. 42).

Besides meeting the formative function of an evaluation system, peer review also helps in fulfilling the summative objective as well. A few specific studies give credibility to the summative portion of the peer evaluation process. As mentioned earlier, traditional teacher evaluation has fallen short in holding teachers accountable, with minimal action taken against ineffective teachers (Loup, Garland, Ellett, & Rugutt, 1996; Tucker, 1997). Murray, Grant, and Swaminathan (1997), on the other hand, pointed out that the peer assistance and mentoring program in Rochester City School District resulted in an average of 10-15% teachers not being rehired because their mentor teachers felt that they did not demonstrate that they could be successful teachers. In addition, some teachers voluntarily elected not to continue teaching as well, often because they are counseled out
by their mentors, while the remaining teachers showed improvement in practice. Goldstein and Noguera (2006) also discovered that districts that implemented peer assistance and review actually retained more new teachers who were competent since they received mentoring and support. For instance, the Columbus Peer Assistance and Review Program which serves 4,800 teachers resulted in 80% of new teachers still working after five years on the job, while in other urban districts without peer review, over 50% of new teachers leave within five years (Bradley, 1998). This point is important to mention because peer review is not just about dismissing incompetent teachers but also about retaining competent educators through peer assistance and appropriate recognition. As Bob Chase, President of National Education Association said, “To categorize peer assistance and review as getting rid of bad teachers . . . is a gross misrepresentation of what it’s all about” (Bradley, 1998, p. 4). One therefore begins to see that the quality of teacher performance becomes the focus with peer reviews.

Other studies have also shown peer reviews implementing the summative function of an evaluation system. Goldstein’s (2004) study of Rosemont Unified School District in California with 100 schools and 3,000 teachers showed that once the district used peer reviews as a part of their teacher evaluation system, the consulting teachers recommended nonrenewal of participating teachers at “unprecedented rate” (p. 191). Those consulting teachers clearly placed greater accountability on their teachers versus past principals who were the sole evaluator. In fact, by having multiple evaluators assigned to each teacher, it increases the accountability for individual evaluators since they know that their notes will be reviewed by other trained evaluators. Finally, in
another mixed design study that looked at the relationship between teacher evaluation scores and value-added analysis for students’ reading, math, and science test scores, Heneman, Milanowski, Kimball, and Odden (2006) discovered that two districts, Cincinnati and Vaughn, had higher average correlations versus the other two districts studied. After digging deeper for possible explanations, the authors posited that the higher correlations at these two districts are due partially because of their use of multiple evaluators through trained peer teachers. In contrast, the two other sites utilized a single evaluator of either the school principal or an assistant principal. The researchers therefore recommended use of peer evaluators because “having a second evaluator provides expertise, reduces workload, and can help reduce leniency when scores have to be compared and discussed” (Heneman, Milanowski, Kimball, & Odden, 2006, p. 9). Additionally, Donaldson (2010) pointed out an added benefit of peer reviews reducing the time problem administrators’ face in evaluating multiple teachers since peer reviews divide evaluation responsibilities between principals and trained teachers. And since these trained teachers are mainly focused on assisting their peers, they have the time to both conduct high-quality evaluations and provide useful and ongoing feedback.

On the other hand, some studies on peer reviews express certain concerns in utilizing educators within a teacher evaluation system. Lieberman (1998) referred to the higher costs of including peer teachers in the evaluation process. The major costs include salaries and benefits for these evaluators, who understandably tend to be senior teachers who are among the highest paid in the district. Also, these teachers are typically the best educators in the district who are taken away from the classroom to focus on peer reviews. It is therefore safe to assume that peer review costs much more than conventional
evaluation procedures. This added cost is not necessarily a reason to reject it; the real question is whether the additional benefits of a peer review are worth the additional costs. Goldstein (2007) includes benefits such as reducing litigation costs associated with terminating tenured teachers by weeding out weak teachers earlier while also improving retention, which avoids the expenses of recruiting, hiring, and orienting new teachers. Lieberman’s (1998) list includes other benefits such as more accurate evaluations, collaborative spirit, higher morale, and increased time for principals to perform other tasks. In actuality, many of these benefits are difficult to quantify. However, with any evaluation component, the cost-benefit analysis should nonetheless be taken into account by individual districts when considering adding an instrument such as peer review into their teacher evaluation system.

Another set of concerns with peer review relate to teacher unions. Bernstein (2004) commented that teacher unions expressed disapproval for teacher evaluating each other because this task is not a part of their job. It can also places teachers against one another. Goldstein (2007) answered these concerns by pointing to the professional role of teacher who takes on a more active role in order to maintain high standards within their profession. In other words, as professionals, teachers should vest themselves with authority and responsibility for the quality of their practice. As Bernstein (2004) wrote, “‘Peer review recognizes a legitimate role for teachers in establishing and enforcing standards in their own occupation’” (p. 82). Furthermore, this professionalism also entails characteristics of “shared knowledge base and a concern for client welfare” (Goldstein, 2007, p. 481). And as teachers take on a greater role to improve their profession, it also transforms schools into a more collaborative environment where
teachers assist one another and teachers and administrators become allies instead of adversaries in the evaluation process (Hertling, 1999).

Another study by Simon & Eby (2003) suggested that by including teachers as evaluators, it can lead to personality incompatibility and power-mongering evaluators who create circumstances in which mentors are resentful of the mentee. To answer this concern, it should be pointed out that anytime individuals are involved in evaluating others, these types of concern exist. Then again, by involving more individuals in the evaluation process, such as the principal and peer teachers, these situations are less likely to take place. Plus, programs such as the Toledo Assistance and Review include a board of review comprised of five union officials and four administrators who ultimately decide the teacher’s status (Toch & Rothman, 2008). These types of programs bring greater credibility to the evaluation process since evaluators must justify their evaluations to a panel of trained professionals. By including union representatives in the review board, it potentially reduces union opposition to teacher dismissals as well (Goldstein, 2004). As Toch & Rothman (2008) wrote, “another way to counter the limited, subjective nature of many conventional evaluations is to subject teachers to multiple evaluations by multiple evaluators” (p. 7).

After looking at concerns in using peer reviews, a few other benefits of this instrument should also be mentioned. Yon, Burnap, and Kohut (2002) explained an additional advantage of peer teachers being able to conduct evaluations at a deeper and more specific content level, which administrators may not be able to provide. Therefore, peer reviews provide greater credibility when combined with other evaluation instruments such as principal observations. In fact, teachers place high regard for
evaluators with extensive knowledge of the content so that they are able to more effectively evaluate and also provide suggestions for improvement (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). As a specific example, the District of Columbia Public Schools included within their evaluation system a master educator because from “over 50 focus groups, DCPS teachers consistently said they wanted an objective, expert teacher, who was familiar with their content area, to be a part of the assessment process” (District, 2009, p. 14). Kimball (2002) reminded his readers that certain perceptions of feedback and fairness are associated with the acceptance and success of the evaluation process by teachers; these perceptions include “perceived evaluator credibility, quality of ideas, depth of information, persuasiveness of rationale for suggested changes, usefulness of suggestions, trustworthiness of evaluator, perceived relationship with evaluator, and perception of evaluator capacity to demonstrate needed changes” (p. 244-245). All these attributes are more likely to be met when peer teacher are also involved in the evaluation process.

Nonetheless, as mentioned earlier, the final evaluation decision should not be placed on the peer reviewer because the function of peer teachers is to provide another component of valuable feedback that gives a more complete picture of teacher practice. Evaluation judgments should be made by others. For instance, Goldstein’s (2004) study indicated that teachers felt that a properly trained administrator should make final judgments after looking at all data sources. Or as in the case of the aforementioned Toledo district, the final decision is made by a board of review (Toch & Rothman, 2008). Essentially, by combining other stakeholders with peer evaluators, the evaluation system gains reliability and credibility through multiple data sources and multiple evaluators.
In summary, peer review programs encourage mentorship, collaboration, and increased professionalism among teachers (Davis, Pool, & Mits-Cash 2000; Milanowski, 2006). By incorporating this component, it gives a more complete and holistic picture of the teacher being evaluated since it brings in another competent participant into the overall multidimensional teacher evaluation system.

Other Instruments

The most common instruments used within a multidimensional teacher evaluation system were described so far. But some researchers and school districts have written about other less widely used components that have the potential to provide additional data sources towards accessing teacher practice. Some of these instruments can be used in conjunction with or serve to complement previously mentioned components. These instruments include teacher self-reflection, analysis of lesson plans and classroom artifacts, and parent feedback.

Teacher self-reflection. In ancient Athens, Greece, carved into the lintel at the Temple of Apollo was written: gnothi seauton or translated to English: know thyself (Hart, 1985). A part of this injunction by philosophers of the past is the call for self-reflection. And in the context of teaching, reflection involves the process where teachers analyze their own practice retrospectively.

At the Center for Research on Accountability and Teacher Evaluation at Western Michigan University, Airasian and Gullickson (1997) listed four steps involved when teachers conduct self-evaluation: problem identification, information gathering, reflection and decision making, and application and change. These reflections can take place in numerous circumstances such as during professional conversations with peer teachers or
administrators or when developing portfolios. Therefore, as mentioned earlier, this component can be used in combination with other evaluation instruments.

Self-reflection can also be a separate instrument given to teachers through instructional logs, feedback from surveys, or interviews in order to measure and report what they are doing in the classroom. By requiring reflection as a part of the evaluation process, it may encourage teachers to continue to develop their practice (Airasian & Gullickson, 1994). After all, teachers themselves have knowledge of their own abilities. But Danielson and McGreal (2000) wrote that many teachers do not take time for self-reflection in a systematic manner unless they were required to do so; therefore, when teachers pause and take the time to reflect on their practice in a systematic manner, “they are richly rewarded” (p. 48).

The extent and level of these self-reflections is broad and as a result depends on its intended purpose and use. For instance, Cohen & Hill (2000) wrote about the basic practice of using self-reports so that evaluators can initially ensure that teachers understand and reflect upon established teaching and curriculum standards. At a greater level of use, these self-report can take the form of instructional logs where teachers keep an ongoing detailed record of teaching practices. These logs are typically highly structured with specific information and agreed-upon set of standards related to instructional practice and curriculum.

One of the main problems with self-reports is the issue of validity. These reflections are susceptible to bias where teachers intentionally or unintentionally misrepresent their own practice (Airasian & Gullickson, 1997). Therefore, if using teacher self-reflections, Ball and Rowan (2004) recommended also gathering multiple
data sources while also ensuring that teachers and evaluators are able to consistently interpret the requirements of the instrument.

**Analysis of lesson plans and classroom artifacts.** Another method to collect data when evaluating teachers is through the collection and analysis of teacher lesson plans and other classroom artifacts. This component is different from teacher portfolio because the artifacts are collected by the evaluator and not the teacher. Analyzing lesson plans, for instance, can provide for the evaluator a gateway into the teacher’s preparation of content. Furthermore, these lesson plans and other artifacts can provide a hint of teacher’s thinking before and after the completion of a lesson. Researchers recommend using specific rubrics to evaluate these artifacts (Denner, Salzman & Bangert, 2001). These rubrics help evaluators access the extent to which teachers plan and carry out their instruction. Stronge (2007) specifically lists what properly prepared lesson plans known to improve student outcomes contain; his list includes: link student learning objectives with teaching activities, describe teaching practices to maintain students’ attention, align student learning objectives with the district and state standards, and accommodate students with special needs.

Other classroom artifacts beside the lesson plan include teacher assignments, tests, teacher scoring rubrics, and any other artifact that helps to determine the quality of teacher preparation and instruction. One study by Clare & Aschbacher (2001) looked at the relation between teacher assignment quality and quality of student work across four elementary and four middle schools. To analyze this relationship, the researchers connected different classroom assignment rating scales with students’ writing rating scales so that they could see how different aspects of teacher assignments impact student...
achievement. The study found that the quality of classroom assignments were statistically significant with the quality of student work, which provided additional evidence for incorporating teaching assignments as an gauge of classroom practice. On the other hand, another study by Matsumura, Garnier, Pascal, & Valdes (2002) in a Los Angeles school district showed that the level of agreement between raters of teacher assignments varied considerably. These same researchers later discovered that two of the expert raters had the highest level of agreement. The study therefore attests to the importance of training so that expert evaluators can consistently use the research-informed scoring rubric for rating student work samples. Furthermore, the evaluators may need to have knowledge of the grade-level and subject matter being evaluated. Both studies also emphasized that at least three assignments with more than one sample each were needed for stable estimates of teacher quality.

The advantages of including classroom artifacts include its practicality since it’s already been created and the fact that it does not place more than needed burdens on teachers (Borko, Stecher, Alonzo, Moncure, & McClam, 2005). Unlike teacher portfolios, teachers do not play a primary role in putting together and analyzing the artifacts, though the burden does fall on the evaluator. Furthermore, tools are available to help evaluate these artifacts; for instance, David Allen and Joe McDonald (2003) from the Coalition of Essential Schools developed a tuning protocol with guidelines, norms, and recommended steps to reflect on teacher and student work. On the other hand, examining lesson plans and other artifacts alone provide only one dimensional of teacher practice. After all, a lesson plan on paper may be different from what actually happens in the classroom. But when classroom artifacts are combined with other data sources, it can
provide a greater perspective of teacher preparation, modification, and practice. As Danielson and McGreal (2000) wrote, “Artifacts, combined with classroom observations, enable an evaluator to witness a teacher’s plans coming to life for students” (p. 49).

**Parent feedback.** With the understanding that evaluating teaching is a complex task that involves examining several perspectives and angles, researchers have argued for incorporating greater number of players in the evaluation process. Some have referred to this idea as the 360 degree feedback (Dyer, 2001; Wilkerson, Manatt, Rogers & Maughan, 2000). This view has led to the development of parent feedback surveys as another instrument for collecting data to evaluate teacher effectiveness (Stronge & Ostrander, 2006). First of all, parent feedback helps in incorporating parents further into schools. And parent involvement has shown positive and significant effects on student factors such as grades and attendance (Astone & McLanahan, 1991; Fan & Chen, 2001; Sheldon, 2003). Stronge and Ostrander (2006) therefore contend, “One vital aspect of enhancing parental involvement in education is to invite parents into the educational enterprise as partners in decision making” (p. 130). Furthermore, these parents provide valuable perspective and insight into teacher performance as well. As an example, Epstein (1985) conducted a quantitative study using multiple regression analysis of parent and principal feedbacks of teachers from 11 school districts in Maryland. The study concluded that parents and principals highlight distinct aspects of teaching in judging teachers' merit, with principals looking at situational factors versus parents who are influenced by teacher connection with families and the quality of classroom life their children experience. In other words, the study revealed that principals and parents emphasized different aspects of teaching. More specifically, the researcher discovered
that teachers earn higher ratings from parents when teachers incorporate parent involvement activities more often, send more communication home, and maintain good classroom discipline. Parent feedback therefore provides valuable information on different teacher skills and abilities that principals may not necessarily see. Furthermore, another study by Faucette, Ball, and Ostrander (1994) revealed that parents appreciate being involved in the evaluation process and teachers generally accepted parent feedbacks, especially finding valuable the comment section of the survey.

On the other hand, Danielson and McGreal (2000) express caution when designing these feedback surveys “so that such information is valid and not compromised by personality differences, favoritism, or other irrelevant matters” (p. 51). In addition, the surveys should include questions that parents can answer, such as teacher accessibility, while also ensuring that it is not excessively detailed. And finally, parents should be assured of confidentiality so that the evaluators can respond with openness and honesty without the fear of any degree of reprisal (Dyer, 2001).

Another concern with using parent surveys is the lack of parent involvement, especially from low-income and minority parents (Desimone, 1999). In fact, studies have shown that White middle-class parents are usually the most visibly active in public schools (Lee & Bowen, 2006; Manz, Fantuzzo, & Power, 2004). Therefore, issues regarding small or disproportionate number of completed parent surveys should be considered. Schools should ensure that they are getting a representative view of the majority of parents instead of the opinion of a select few. In fact, caution should be used when placing too much trust in the results of a single data set such as parent surveys. For instance, Danielson and McGreal (2000) mentioned that because parent feedbacks are
based on perceptions, evaluators should not consider them as “entirely reliable sources of evidence” (p. 51). Instead, they are to be used as a supplement to other indicators of teacher performance. Epstein (1985) concluded his study of parent surveys with the disclaimer: “Because there is not a single set of skills that perfectly define effective teaching, measures of many aspects of teaching by multiple judges are likely to yield the fairest and most comprehensive evaluation of teachers” (p. 8). Parent feedback is one other instrument that can provide an additional perspective of teacher performance.

**Conclusion**

A multidimensional teacher evaluation system can incorporate multiple components such as student achievement data, teacher portfolio, and principal, peer, and student evaluations. Inputs are balanced with student learning outcomes. These various instruments likewise make use of formative and summative evaluations in order to hold teachers accountable while also improving their practice. In the end, Peterson, Stevens, and Ponzio (1998) concluded from their research study that they favor the use of multiple data sources within a teacher evaluation system for the following reasons:

- No single data source is sufficient for all purposes. Excellence in teaching comes in a variety of configurations and areas of performance.
- Good teachers are good for different reasons. The kinds of information that are most helpful in understanding quality vary from one individual teacher to another.
- Not all data are available in one setting.
- The contexts of teaching vary considerably.
Essentially, the use of multiple data sources with multiple evaluators ensures that various aspects of the complex activity of teaching are encompassed within the evaluation system.

On the other hand, Peterson (2000) pointed out that a major disadvantage of including multiple data sources in an evaluation system is the subsequent requirement of more time from evaluators and teachers. So choices should be made on which instruments to use. Just as there is no single set of data sources that fit all teachers, there is also no single recommendation on which instruments to use. Peterson (2000) with Danielson and McGreal (2000) therefore recommended that the choice of instruments should come from either the district or school buildings or even at the individual teacher level through differentiated procedures that vary according to teaching experience and particular area of professional growth. Furthermore, the design of a teacher evaluation system should also incorporate measures that take into account the different contexts across grades, content areas, student subgroups, varied student populations between schools, school culture, and local circumstances (Campbell, Kyriakides, Muijs, & Robinson, 2003). Essentially, “Evaluation developers need to consider which activities will be suitable for documenting the different evaluative criteria” (Danielson & McGreal, 2000, p. 57). In other words, it is not necessary to develop a universal evaluation system; instead, the ideal instruments would depend upon the needs and purposes of the district, school, or teacher. Plus, schools can also expand slowly in incorporating additional instruments into their evaluation system. Ultimately, when teachers and administrators are involved in the process of choosing and developing the evaluation system together,
they are more likely to accept and buy-in to the process as well (Darling-Hammond, 1983; Peterson, 2000).

Finally, Danielson (2007) reminded her readers that a framework for teaching with the use of multiple instruments cannot by itself ensure an effective teacher evaluation system. It will depend on how the system is used. Factors such as “the professional culture of the school and the district, the degree of respect among and between teachers and administrators, and the commitment of all educators to ongoing improvement of practice” all factor into the effectiveness and success of the multidimensional teacher evaluation system (p. vii).

In summary, we know that instructional expertise of teachers is at the heart of the learning enterprise (Darling-Hammond, 2000; Wang, Haertel, & Walberg, 1993). Teacher evaluation should therefore be an important focus in improving education because “without high quality evaluation systems, we cannot know if we have high quality teachers” (Stronge & Tucker, 2003, p. 3). After all, the center of education is the link between teaching and learning, and this relationship is strongest when effective teachers are educating students. In other words, we know that teacher evaluation is a vital mechanism for improving teaching and learning (Beerens, 2000; Borman & Kimball, 2005; Danielson & McGreal, 2000; Darling-Hammond, 2006a; Peterson, 2000). And when studying teacher evaluation systems, multiple data sources and evaluators provide a more comprehensive understanding and perspective of what teachers do.

Additionally, when studying multidimensional teacher evaluation systems, we know that it is important to also have a common definition and understanding of the complex activity of teaching. Danielson’s Framework for Teaching (2007) helps to
describe various components of a teacher’s daily work and underlying assumptions of the responsibilities of teaching. Moreover, the Danielson framework has one of the greatest longevity in the field and is also one of the most widely used framework (Brandt, Mathers, Oliva, Brown-Sims, & Hess, 2007; Charlotte, 2011). Irrespective of subject area or grade-level, the framework addresses the various characteristics of teaching through the use of four domains: planning and preparation, the classroom environment, instruction, and professional responsibilities. So we know that the framework’s key purpose is to serve as a guide for both shared understanding and common language of effective teaching, which is essential when studying teacher evaluation systems.

On the other hand, though we know a lot about the makeup and the “what” of a multidimensional teacher evaluation system, there is a significant lack and need for literature related to the evaluation process and procedure, or the “how-to” of these systems. For instance, the Brookings Institute began its report on teacher evaluation by stating: “U.S. public schools are in the early stages of a revolution in how they go about evaluating teachers” (Glazerman et al., 2011, p. 1). A different study by the National Council on Teacher Quality expressed the same sentiment when they wrote, “It is important to note that the development of teacher evaluation systems is a moving target, and we are in a period of rapid change” (National, 2011, p. 1). Due to the novelty in reforming teacher evaluation systems, the previously mentioned Brookings Institute report admitted an “immaturity of the knowledge base on the design of teacher evaluation systems” (Glazerman et al., 2011, p. 1). Another study by the National Comprehensive Center for Teacher Quality also noted “the dearth of available research-based methods and models of comprehensive teacher evaluation” (Goe, Holdhelde, & Miller, 2011, p. 5).
This deficiency and need is especially evident in Michigan with the state’s recent legislative mandate that their evaluation system must include multiple data as measures of educator effectiveness (S. Res. 1509). In fact, the National Council of Teacher Quality’s report (2010) specifically looked for improvements on “Michigan policies that need to better connect to teacher effectiveness” (p. 7). Furthermore, recent reports from both the Education Alliance of Michigan (2011) and the Michigan Association of School Personnel Administrators (2010) acknowledged that there are few multidimensional teacher evaluation models that currently exist in Michigan schools. As Danielson and McGreal (2000) wrote, “the ‘how’ of a teacher evaluation system is at least as important as the ‘what’ of the system” (p. 62). This study will therefore attempt to partially fill this void and deficiency by examining what strategies and processes are used by schools in Michigan that are attempting to build research-supported multidimensional teacher evaluation systems.

Such a study is particularly important in Michigan because it will assist state policy makers, local districts, and individual schools in assessing current evaluation practices towards meeting recent Michigan standards of a “rigorous, transparent, and fair performance evaluation system” (S. Res. 981, 2009). It must also be mentioned that though innovations in multidimensional teacher evaluation systems are beginning in Michigan schools, often the knowledge of these processes and practices is limited to the school site. This study will subsequently also help in disseminating information on how Michigan schools are moving toward developing multidimensional evaluation practices and systems.
Even more importantly, a study focused on developing and improving teacher evaluation systems is vital in improving teacher practice. After all, teacher evaluation matters because teaching matters since research has shown that teacher quality is one of the most important predictors of student achievement (Darling-Hammond, 1997b). Therefore, by investing in improving teacher evaluation, student achievement is ultimately impacted in a positive way.
CHAPTER III

METHODOLOGY

Research Design

Yin (2003b) wrote that all types of empirical research, implicitly or explicitly, make use of a research design. This research design functions in connecting the data of the study to both its initial research question and its final conclusions. In other words, the research design “is a logical plan for getting from here to there, where here is defined as the initial set of questions to be answered, and there is some set of conclusions (answers) about these questions” (Yin, 2003, p. 20). Alternatively, Hancock and Algozzine (2006) viewed the research design as a framework that provides a road map. This design or map looks at various factors such as whom to study and how best to acquire, then analyze the information. Essentially, the research design provides the structure or framework for the research study.

There are mainly two types of research designs. These two methodologies, quantitative and qualitative research, present two varying approaches to the understanding of social science and the study of human beings. Travers (2001) contrasted the two traditions through the writings of Emile Durkheim and Max Weber. Durkheim’s ideas fit the perspective that quantitative methods are seen as more ‘scientific’ since, like natural science, it makes causal connections between variables. On the other hand, Weber’s writings opposed the viewpoint that social science should emulate natural science. After all, human beings are different from inanimate objects because they think and experience emotion. And Weber believed that this additional dimension of the social world should be examined. In fact, Donmoyer (1990) argued that
researchers in education are also concerned with individuals and not only aggregates; “for them, questions about meaning and perspective are central and ongoing” (p. 66).

Therefore, the second approach of qualitative research focuses on making sense of how people understand their own actions, with the key philosophical assumption that “reality is constructed by individuals interacting with their social worlds” (Merriam, 1998, p. 6).

From these two perspectives of conducting research, this study will focus on the qualitative methodology. According to Creswell (2009), qualitative research studies “support a way of looking at research that honors an inductive style, a focus on individual meaning, and the importance of rendering the complexity of a situation” (p. 4). This type of approach stands in contrast to quantitative studies that focus on explanation and control, where the researcher takes on an impersonal role (Stake, 1995). Given that this research study mainly takes place within natural settings of Michigan schools and involves reporting multiple perspectives on teacher evaluation models, a qualitative approach is appropriate. After all, “‘qualitative researchers study things in their natural setting, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them’” (Gall, Gall, & Borg, 2007, p. 31).

Specifically, this study will examine schools attempting to develop a multidimensional teacher evaluation system through the use of a case study approach. Merriam (1998) wrote that the “single most defining characteristic of case study research lies in delimiting the object of study, the case” (p. 27). It is a bounded system or unit with boundaries of what is to be studied. Merriam goes on to say that the case could be a program or system. Moreover, Creswell (2009) 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. 13). In fact, Merriam (1998) later pointed out, “Case study is a particularly suitable design if you are interested in process” (p. 33). For these reasons, the case study strategy of inquiry matches the overarching research question of this study, which examine the process of building an evaluation system by asking: What strategies and processes are used by schools in Michigan that are attempting to build research-supported multidimensional teacher evaluation systems based on Charlotte Danielson’s *Frameworks for Teaching*. From this overarching question, there are also four primary sub-questions:

1. How are schools addressing each of Danielson’s four domains of teacher practice as they begin to make their evaluation practices and systems more multidimensional and transformational (growth-oriented)?

2. What resources are utilized to implement a multidimensional teacher evaluation system?

3. What challenges do schools face in implementing a multidimensional teacher evaluation system and how are they addressed?

4. What evidence(s) exists that the teacher evaluation system has or is resulting in positive educational outcomes?

These sub-questions go deeper into examining the evaluation process within the setting of two Michigan schools by focusing on resources, challenges, and evidence of growth that these schools exhibit when developing a multidimensional model. Marshall and Rossman (2011) wrote, “Case studies take the reader into the setting with vividness and detail not typically present in more analytic reporting formats” (p. 267). The research sub-questions calls for such details about the setting of the two schools that will be
studied, particularly looking at the schools’ resources, challenges, and evidence of growth. Essentially, this study is focused on providing an in-depth analysis of the evaluation process and system, which meets the fundamental characteristics of a case study inquiry (Creswell, 2009; Merriam, 1998). Furthermore, a case study approach allows the unfolding of the implementation of the evaluation system, offering the contractual perspective and the perspectives of those who planned and administered it and also the perspective of those who are evaluated under it. Hence, the model used in this study is a qualitative approach through a case study design that examines the evaluation process within two Michigan schools that will ultimately address both the study’s purpose and research questions.

**Sampling, Subjects, Access and Setting**

I will first need to identify schools in Michigan that utilize Charlotte Danielson’s *Framework for Teaching* within their multidimensional teacher evaluation system. Yin (2003a) recommended in his section for case screening the following: “contacting numerous individuals in the field and consulting available reports and literature” (p. 13). Therefore, I will make contact with the Office of Accountability, Research, and Evaluation of the Michigan Department of Education to help identify potential corresponding school districts. Additionally, referrals from individuals involved in the teacher evaluation movement will be utilized as well. Once potential districts are identified, I will then select schools after contacting them and reviewing their website and other documents related to their teacher evaluation practices. Information regarding the schools’ demographic data will also be collected through the Michigan Department of Education website.
It must be mentioned that reports such as The Education Alliance of Michigan (2011) acknowledged that there are few comprehensive teacher evaluation models in Michigan schools. Consequently, two schools will be selected based upon the following four criterion: their demonstration of research-related multidimensional teacher evaluation practice based on Danielson’s *Framework for Teaching*, use of student growth data within the evaluation system, at least one year of implementation, and willingness to participate in the study. Selecting multiple schools is actually recommended by Yin (2003b) over a single-case design because single-case designs “are vulnerable if only because you will have put ‘all your eggs in one basket’” (p. 53). On the other hand, selection is limited to two school sites because “the study of multiple cases reduces the total attention that can be given to any one of them, and thus serves to weaken rather than to strengthen the study” (Gall, Gall, & Borg, 2007, p. 457). By limiting the selection to two schools, I am therefore able to study each case in greater depth. Hence, by limiting the selection to two schools that meet the above criteria, I am able to study each case in greater depth while also helping to avoid criticisms around a single-case study.

Location and feasibility of conducting a research study at these sites will be considered so that the schools can be readily accessed and data easily collected. This type of sampling strategy meets the purpose of a convenience sampling where savings of time, money, and effort for the researcher is important (Creswell, 2007; Merriam, 1998). And in the end, finding schools that are open to conducting the research study is a vital factor; in fact, Stake (1995) writes, “opportunity to learn is of primary importance” (p. 6).

After obtaining permission from a human subjects review board, I will then gain access to these school sites by first contacting the schools’ principals through personal
visits. Gall, Gall, and Borg (2007) brings up an important strategy in gaining entry: “deciding how to phrase your request (e.g., focusing on the site’s opportunity to contribute to research or on personal benefits to site participants)” (p. 458).

Consequently, I will provide a recruitment letter to the principals explaining the purpose and benefits of this study. Benefits from this study such as dissemination of their practices to other schools and the understanding of issues in the implementation process from multiple perspectives will be emphasized in the letter and during my personal visits with the principals. Principals will also have the added benefit of a full report of their school’s evaluation process at the end of the study. Furthermore, ethical concerns and the respect for persons will be clarified in the letter and during my visits such that I, the researcher, will respect all participants’ privacy, anonymity, and their right to participate or not (Marshall & Rossman, 2011). Lastly, I will obtain written permission from the superintendent of the district before proceeding with any part of the study.

Once two schools that meet the aforementioned criteria accept participation in writing, I will take the following steps to identify participant pool. Prior to visiting the school site, I will conduct a pre-site phone interview with the principal to inform the school of their participation in the study, to collect general information regarding their evaluation process, and to schedule site visits. Principals will also be asked to provide copies of school documents, such as protocols that were used in their evaluation program implementation, and to identify individuals involved in the implementation of the evaluation practice. These individuals will form the significant portion of my participant pool. This pool will consist of the following in each school: a central office personnel involved in developing the evaluation model, an administrator in charge of evaluating
teachers, a lead teacher or grade-level chair, and a select number of teachers from different levels and/or subject areas. Additionally, if the participating schools utilize student and/or parent feedbacks in their evaluation model, they will also be incorporated into the participant pool. The criteria for this participant pool will be based upon principal recommendations and participants having a leading or direct involvement in teacher evaluation process at their respective schools. This type of approach connects with the concept of purposeful sampling which means that “the inquirer selects individuals and sites for study because they can purposefully inform an understanding of the research problem and central phenomenon in the study” (Creswell, 2007, p. 125). The recruitment of these participants will primarily come from principal reference and my personal visits with each participant explaining the purpose, confidentiality, and benefits of the study as previously mentioned.

**Data Collection Methods, Procedures and Instrumentation**

Creswell (2007) recommended that “case study data collection involves a wide array of procedures as the researcher builds an in-depth picture of the case” (p. 132). Furthermore, Yin (2003a) wrote that the richness of the context of case studies “means that the study cannot rely on a single data collection method but will likely need to use multiple sources of evidence” (p. 4). To help sort the numerous types of data, Rapley (2007) divided them simply into two categories: “data that you have to generate and data that already exists” (Rapley, 2007, p. 8). This study will specifically involve gathering multiple forms of data through document analysis, interviews, survey, and observations. Document analysis originates from data that already exists while interviews, survey, and observations come from data that is generated.
Multiple school visits will first be scheduled with the participants once they are identified. During these visits, the following procedures will be utilized for data collection: analysis of documents related to the evaluation process, interviews with central office personnel, administrator, and a lead teacher who are involved in the formation or implementation of the evaluation process, focus group interviews with select number of teachers, observations of both the implementation of the evaluation model and teacher evaluation professional development meeting, and finally an online survey made available to all teachers. My research questions will guide all these data collection processes. In other words, information from each data collection instrument will address one or more of the research questions. Marshall and Rossman (2011) suggests accordingly by writing that the researcher “should consider how the selection of any particular method will inform the research questions, thereby extending and deepening knowledge on the topic” (p. 168).

**Document Analysis**

Before conducting interviews and observations, I will first complete document analysis at each site with the permission of the principal. Bowen (2009) defined document analysis as “a systematic procedure for reviewing or evaluating documents—both printed and electronic material” (p. 27). Since documents that fit this terminology are plentiful, I will determine which of these documents will be reviewed and evaluated based upon its relevance to the study’s research problem and purpose. These documents will initially provide background and historical information related to the school’s teacher evaluation system while also presenting the contractual point of view of the evaluation process as well. Marshall and Rossman (2011) also pointed out that “the
analysis of documents is potentially quite rich in portraying the values and beliefs of participants in the setting” (p. 160). Furthermore, documents provide supplementary research data on the subject matter that serves as valuable additions to the knowledge base (Bowen, 2009). Therefore, important documents related to the implementation of the school’s teacher evaluation model will first be collected by looking at evaluation handbooks, teacher contract, performance criteria and rubrics, and timelines for implementation, as well as sample teacher improvement plans and meeting minutes, and any other pertinent information related to the evaluation process.

There are a few advantages in using the instrument of document analysis. Bowen (2009) listed the following that apply to this study:

- Efficient method that requires less time-consumption
- Cost-effective since the data contained in documents already exist
- Lack of obtrusiveness
- Documents are stable, where researcher’s presence does not alter what is studied
- Documents provide broad coverage across long span of time, events, and settings

(p. 31).

At the same time, it is important that I continually access the authenticity and representativeness of the documents collected while I also consider the original purpose and intended audience of these documents as well (Rapley, 2007).

The participant pool will help in identifying potential documents to use, as well as the school secretary. The documents will be recorded through the use of a document summary form where “the researcher writes a brief summary of each document that has been examined, noting the type of document, its uses, a summary of its contents, and
ideas about other documents that should be obtained and studied” (Gall, Gall, & Borg, 2007, p. 463). And finally, Yin (2003b) reminded his readers, “For case studies, the most important use of documents is to corroborate and augment evidence from other sources” (p. 87). Thus, information obtained from these documents will help guide the proceeding individual interview, survey, focus group, and observation.

**Individual Interview**

Because of their important role in the teacher evaluation process, interviews will be conducted with a central office personnel, administrator, and select teachers. Kvale and Brinkmann (2009) defined an interview as “an inter-view where knowledge is constructed in the inter-action between the interview and the interviewee” (p. 2). For this study, it is important to know how people involved in the teacher evaluation process understand and make sense of the process. One way of accomplishing this objective is by talking to them; after all, conversation is a basic form of human interaction. “Qualitative researchers rely quite extensively on in-depth interviewing” (Marshall & Rossman, 20011, p. 142) because “most case studies are about human affairs” (Yin, 2003b, p. 92). Essentially, interviews help to report and interpret these human affairs through the interviewee, whereby providing valuable insight into the specific case (Yin, 2003b).

Each of these in-depth interviews will be tape recorded after obtaining consent from participants and also aided by an interview protocol with questions that serve as a guide to the interview process. Therefore these interviews will make use of semi-structured interview protocol with more structured probes in order to obtain perspectives of the evaluation process. Again, these questions will be based upon the research questions for this study and also information obtained from initial document analysis.
Stake (1995) noted that the purpose of these questions “is not to get simple yes and no answers but description of an episode, a linkage, an explanation” (p. 65). Hancock and Algozzine (2006) further pointed out that these interviews should be semi-structured, where researchers ask predetermined but flexible worded questions such that the interviewer is able to respond to the situation at hand. Follow-up questions will also be considered beforehand in order to probe more deeply into the issues. Lastly, I will make use of probes, recommended by Rubin and Rubin (2005), in order to manage the conversation through verbal and nonverbal techniques when necessary. These probes include continuation, elaboration, attention, clarification, steering, sequence, and evidence probes, which are essentially all designed to “manage the conversation by regulating the length of answers and degree of detail, clarifying unclear sentences or phrases, filling in missing steps, and keeping the conversation on topic” (Rubin & Rubin, 2005, p. 164). Conversely, I will at the same time practice and attempt to follow Seidman’s (2006) suggestion that “Listening is the most important skill in interviewing” (p. 78). It is a type of active listening that requires genuine concentration and interest in the interviewee while also remaining aware of how to move the interview forward. Quintessentially, the interview is a “‘conversation with a purpose’” (May, 2002, p. 226). Finally, before visiting the sites for interviews, I will initially conduct pilot testing to practice and reform the interview protocol to make it as natural as possible.

Before detailing the makeup of the individual interviews at each site, it is first important to note certain common characteristics. In fact, the three characteristics of varying expertise and perspectives, first-hand experience, and depth of understanding fit the term of responsive interviewing, coined by Rubin and Rubin (2005). First of all, the
The purpose of these interviews is to gain a variety of perspectives of the evaluation process. Rubin and Rubin (2005) wrote that responsive interviews “recognize that each conversational partner has a distinct set of experiences, a different construction of the meaning of those experiences, and different areas of expertise” (p. 34). These varying areas of expertise include the perspectives of a central office personnel, administrator, lead teacher, and classroom teachers from different grades and subjects. Additionally, these subjects are chosen because they have first-hand experience in the evaluation process, which according to Rubin and Rubin “is critical in making your results convincing” (p. 65). And lastly, an additional goal of these interviews is to gain greater depth of understanding. By combining these interviews with document analysis and observations, I hope to consequently gain a deeper understanding of the evaluation process at the two schools.

The study will conduct interviews in the following manner. The individual interviews with a central office personnel and the administrator will last for approximately an hour each. The interview protocol in Appendix A will be used to garner information about evaluation implementation details, theory of action, goals, evidence of teacher and student growth, challenges, lessons learned, and resources utilized. This type of interview is known as the interview guide or topical approach where the interview is scheduled and “the interviewer comes prepared with a list of topics or questions” (Marshall & Rossman, 2011, p. 144). My purpose of the initial interview is to gather general information regarding program implementation and also to identify additional artifacts, as well as to understand the implementation of teacher evaluation from a central office and administrative perspective.
In addition to the individual interview with a central office personnel and the administrator, another interview will be conducted with a lead teacher or grade-level chair since they serve important supporting roles in the evaluation process. The protocol for this interview will be similar to the administrative interviews, which will also last for approximately an hour, but with the purpose of gaining an understanding of the evaluation process from a non-administrative teacher evaluator.

**Focus Group**

In addition to individual interviews, select teachers from each school site will then be consulted. Since a group of teachers consisting of three to four members from each site will be interviewed to obtain their perspectives of the evaluation process, a focus group approach is appropriate. Krueger and Casey (2000) defined focus group as “a carefully planned series of discussions designed to obtain perception on a defined area of interest in a permissive, nonthreatening environment” (p. 5). Litosseliti (2003) listed the following strengths in using focus groups that also connect to this study’s purpose:

- Discovering new information and consolidating old knowledge
- Obtaining a number of different perspectives on the same topic, in participants’ own words
- Gathering information on participants’ views, attitudes, beliefs, responses, motivations, and perceptions on a topic; ‘why’ people think or feel the way they do
- Brainstorming and generating new ideas, with participants discussing different angles of a problem, and possibly helping to identify solutions (p. 18).
Creswell (2007) also wrote that focus groups are advantageous “when interviewees are similar and cooperative with each other, when time to collect information is limited, and when individuals interviewed one-on-one may be hesitant to provide information” (p. 133). And lastly, Marshall and Rossman (2011) pointed out that “in program design and evaluation, focus groups are especially useful” (p. 149). Therefore, focus groups provide valuable information for this study through the distinctive use of participant interactions.

The structure of these focus groups will be important. It should be noted that the group size will be small with three to four participants since more in-depth insights are desirable and the members also have direct and lengthy experience in the topic of discussion (Morgan, 1997). Litosseliti (2003) also recommended against participant homogeneity if diverse opinions and experiences are preferred. Seidman (2006) refers to this strategy as “maximum variation sampling” where the range of people interviewed reflects a fair representation of the population (p. 52). For this reason, the focus group will include teachers from different levels and subject areas. Additionally, a vital role when carrying out focus group sessions is the moderator. Their task is to maintain the group’s focus with key questions examined together while also ensuring that participants neither dominate nor drift away from the topic of discussion. The moderator is able to fulfill these function if they have “good personal, interpersonal, communication and managing skills” who appears “neutral’, opinion-free and non-judgmental” while at the same time holding a delicate balance between being “confident and in control” and also “flexible and adaptable” (Litosseliti, 2003, p. 42).

Researchers provide other important guidelines when making use of focus groups. Morgan (1997) suggested starting the session with an introduction and then a discussion-
starter question that is easily answerable by all participants. Thereafter, an interview protocol will focus on questions that address procedures, strengths, weaknesses, and resources utilized throughout the evaluation process in the particular school. During these discussions, I will utilize the following two techniques offered by Krueger and Casey (2000): the five second-pause to encourage comments, and the second technique of probing, which is also used in individual interviews, in order to request additional information through questions such as “Would you explain further?” or “Is there anything else?”. Moreover, I will make use of these probing techniques to elicit differing points of view through questions such as “Does anyone see it differently?” or “Are there other points of views?” (Krueger & Casey, 2000, p. 110). And lastly, the session will end with the adjourning stage to allow for reflection by group members (Litosseliti, 2003).

Similar to the individual interviews, these focus group sessions will make use of semi-structured interview protocol with more structured probes in order to obtain teacher perspectives of the evaluation process. The protocol in Appendix B includes the aforementioned recommendations, which will be used as a guide during the focus group sessions. These sessions will last for approximately 30 minutes. It is important to note that when scheduling any type of interviews, I will prioritize my time afterwards for self-reflection; Stake (1995) provided this tip when he wrote, “Perhaps the most important thing is to insist on ample time and space immediately following the interview to prepare the facsimile and interpretive commentary” (p. 66).

Observation

In addition to personal and focus group interviews, observations of both the implementation of the evaluation model and professional development meetings related
to evaluation procedures will also be carried out. Marshall and Rossman (2011) reminded their readers, “Observation is a fundamental and highly important method in all qualitative inquiry” (p. 140). One reason for this is because “observations work the researcher toward greater understanding of the case” (Stake, 1995, p. 60). The purpose of these observations is to provide an account of the implementation process and also interactions between administrators and teachers in encounters related to the teacher evaluation process. I, as the researcher, will serve in the capacity of observer only. This role as observer is important; as Stake (1995) wrote, “For all their intrusion into habitats and personal affairs, qualitative researchers are noninterventionists. They try to see what would have happened had they not been there.” (p. 44). I will therefore try to observe the ordinary with minimal interference so that naturalistic observations come about. Merriam (1998) further offered the advice on what to observe by focusing specifically on the conceptual framework, the problem, and the questions of interest. I will collect data from these observations either on the evaluation implementation observation protocol (Appendix C) or the professional development observation protocol (Appendix D), which will both consist of two sections. The first section will be my descriptive notes that include “portraits of the participants, a reconstruction of dialogue, a description of the physical setting, accounts of particular events, or activities” (Creswell, 2009, p. 181-182) while the other section incorporates my reflective notes. And lastly, the observation framework will be based upon information from my initial document analysis and interviews in order to validate findings.


**Online Survey**

After completing interviews, focus groups and observations, an online survey will finally be given to the staff at each school. Yin (2003) included formal surveys as a type of interview, involving more structured questions. Such a quantitative survey will be utilized in this study to obtain a more comprehensive feedback from teachers. Hancock (2006) wrote that one of the advantages of surveys is that such an instrument “created by the researcher often provide a powerful means by which to collect information pertaining to the researcher’s questions” (p. 52). The online survey administered through Google will make use of 4-point Likert-type scale questions of strongly agree (4) to strongly disagree (1) with the option of “don’t know”. These questions will be focused on exploring the perceptions of teachers regarding the use of their teacher evaluation system. Specifically, the survey items will be based upon the four essential attributes of sound and fair evaluations, according to the Joint Committee on Standards for Educational Evaluation (2009); these attributes include:

1. Propriety standards to ensure legal, ethical, and due regard for the welfare of the evaluatee and those involved in the evaluation.

2. Utility standards to guide evaluations so that they will be informative, timely, and influential.

3. Feasibility standards to ensure that the system is as easy to implement as possible, efficient in their use of time and resources, adequately funded, and viable from a political standpoint.

4. Accuracy standards to determine whether the evaluation has produced sound information and are technically adequate and as complete as possible (p. 6-7).
Essentially, these standards will be utilized to examine the efficacy of the teacher evaluation model at each school. In addition to these items, general questions will lastly be included in the survey regarding professional development and the teacher’s perception of the school’s previous evaluation model. The online survey link will be emailed to the staff at each school, with participation being voluntary and anonymous.

In summary, the research study will involve gathering multiple forms of data through document analysis, interviews, observations and survey. It should also be mentioned that specific instruments used within each school’s evaluation model will likewise be explained in detail as well. Yin (2003b) noted that “the most important advantage presented in using multiple sources of evidence is the development of converging lines of inquiry” (p. 98). Furthermore, multiple data sources provide a more holistic account of the complex process of conducting a multidimensional teacher evaluation model. As an example, Seidman (2006) pointed out the benefit of making use of both observations and interviews; observations provide access to behaviors while interviews place those behaviors in context to help understand it. Furthermore, this case study approach offers multiple perspectives, ranging from the contractual perspective to the perspectives of those who planned and administered it and also the perspective of those who are evaluated under it. Subsequently, with the use of these multiple data sources and perspectives, the findings and conclusions in the case study are more likely to be convincing, accurate, and valid (Patton, 1999; Yin, 2003b).

**Data Collection Phases**

The data collection for this research study was completed in two separate phases. It should first be noted that printed and electronic documents were collected throughout
these phases, providing both background information related to the school’s evaluation system and also supplemental research data that served as valuable additions to the knowledge base. Collected documents included evaluation handbooks, teacher contracts, performance criteria and rubrics, timelines for implementation, sample teacher evaluations and portfolios, and finally meeting minutes from trainings related to the evaluation process. These documents were recorded with a document summary form, as recommended by Gall, Gall, and Borg (2007), which detailed the type, use, and summary of the content.

In the first phase, I conducted interviews with the principal, lead teacher, and central office representative. These interviews used a semi-structured interview protocol with more structured probes based on the study’s research questions. They were audio-taped and field notes were written down. Then it was transcribed verbatim. Soon thereafter, I reviewed the transcribed interviews and field notes to identify major topics from the administrators’ and lead teacher’s point of view. Subsequently, I used the preliminary analysis of these interviews and field notes to develop focus group interview protocol.

In the second phase of data collection, I began with focus group interviews consisting of three to four teachers from different grade levels and subjects. These interviews were audio-taped with field notes written down. It was then transcribed verbatim. Afterwards, the transcribed focus group interviews and field notes were reviewed to identify issue and topics from the teachers’ point of view. Then I conducted classroom and teacher training observations. The protocol for these observations, as recommended by Creswell (2009), included descriptive notes on one side and reflective
notes on the other with demographic information such as place, time and date of the field setting also recorded. Afterwards, I used main topics and issues derived by a preliminary review of interview, focus group, observation and document data as a guide in adapting the online survey. The protocol for this survey was based upon the four essential attributes of sound and fair evaluations according to the Joint Committee on Standards for Educational Evaluation (2009). Adaption was made to it based upon any other subject that needed to be addressed according to the study’s research questions. Such a survey was developed in order to obtain a more comprehensive teacher perspective of the evaluation process. A link to this online survey, administered by Google, was finally provided by email to the general teaching staff, with participation being voluntary and anonymous.

**Data Analysis Processes and Procedures**

While data is collected through document analysis, interviews, and observations, it is important for me to continually reflect about the data and write memos to myself throughout the study. In other words, “qualitative data analysis is conducted concurrently with gathering data, making interpretations, and writing reports” (Creswell, 2009, p. 184). And from these ongoing reflections and various methods of data collection, I will then attempt to organize and make meaning of the data.

This study will approach data analysis with a focus on real-life data. Marshall and Rossman (2011) define this approach as in vivo codes, which are codes that come from the actual data. In other words, this type of data analysis involves segments of text that the researcher lifts out while also trying to stay as close to the data as possible. Using this type of analysis, I will first organize and prepare the data for analysis by
transcribing my various interviews, typing field notes from my observations, and optically scanning documents and taking notes. Once the data is collected and organized, the various phases of data analysis will be detailed in the next chapter. These phases are based upon the overarching approach of inductive analysis using the study’s research questions.

The final step in data analysis is the interpretation of the data. Marshall and Rossman (2011) write, “‘Interpretation means attaching significance to what was found, making sense of the findings, offering explanations, drawing conclusions, extrapolating lessons, making inferences, considering meanings, and otherwise imposing order’” (p. 480). These interpretations will be focused on answering my overarching research questions and sub-questions. Stake (1995) supports this approach because equal attention cannot be given to all data; therefore, “The case and the key issues need to be kept in focus. The search for meaning, the analysis, should roam out and return to those foci over and over.” (p. 84-85). Additionally, these interpretations will be based upon a blend of my analysis of the data combined with information obtained from research literature. I will also identify connection of findings to the literature review with conformations and divergences pointed out. And finally, new questions and issues may be raised from these findings for further study.

**The Researcher**

After four years of teaching at public schools and helping administrators through the evaluation process, my interest in this subject matter has grown with time. Initially, I experienced firsthand the shortcomings of traditional teacher evaluation systems that lacked a focus on developing teachers. Then I started studying about the potential of
multidimensional teacher evaluation systems based on research-proven frameworks for teaching. But the question still remained as to how this model could be developed and implemented in a dynamic and complex school setting. This study helps to address that question. As Foss and Waters (2007) suggests, “Writing a dissertation, like taking a trip, should be fun and interesting” (p. 3). The topic of teacher evaluation certainly meets these conditions since a qualitative case study approach allows me to investigate, like a detective, how an evaluation model is developed in school sites (Merriam, 1998; Yin, 2003b).

With the above personal perspective and interest in the topic of this research study, it is important for me to not allow “biased views to influence the direction of the findings and conclusions” (Yin, 2003b, p. 10). Marshall and Rossman (2011) suggested, “Scrupulous qualitative researchers are on guard from the beginning, having been explicit about their voice, their biases, and how their identities have shaped their research questions” (p. 220). Therefore, throughout the study I will constantly reflect and make mention of my personal perspectives and biases when collecting and analyzing data. For instance, in the observation protocol, one section is devoted to my reflective notes that will include these types of data entry. Creswell (2009) goes further by recommending the following eight strategies to access the accuracy of findings: triangulating different data sources, using member checking with rich and thick description, clarifying researcher bias, presenting negative or discrepant information, spending prolonged time in field, using peer debriefing, and using external auditors. Since this study is also a dissertation, audits and debriefing will primarily occur through my committee of professors that I will meet on an ongoing basis. I will also get together with certain colleagues in my program
to look over and discuss my research study. Regarding triangulation, it will take place primarily through my multiple data collection techniques of document analysis, interviews, and observations. From these various data collection approaches, the study will generate detailed and thick descriptions with my frequent reflections for discrepancies. The written narratives will then be taken back to participants for member checking. In essence, by setting up the study beforehand with these multiple strategies incorporated throughout, I, as the qualitative researcher, will help bring about greater reliability and validity into my study.

**Delimitations**

As Foss and Waters (2007) admit, “every study, to be practical, has some limitations” (p. 237). With qualitative case studies, I must first admit that my data collection and analysis of that data cannot be completely objective. After all, “Our words flow from our own personal experiences, cultures, history, and backgrounds” (Creswell, 2007, p. 230-231). On the other hand, I must also be vigilant in constantly reflecting about and admitting these limitations. Nonetheless, “our writing can only be seen as a discourse, one with tentative conclusions, and one that will be constantly changing and evolving” (Creswell, 2007, p. 230-231). In addition, the intent of this type of research is not to generalize findings to settings outside those under study. As Marshall and Rossman (2011) write, “Qualitative research does not claim to be replicable” (p. 254). Furthermore, Yin (2003b) admits, “case studies, like experiments, are generalizable to theoretical proposition and not to populations or universes” (p. 10). Finally, Stake (1995) specifically pointed out that case studies are primarily focused on understanding the unique case itself; he wrote, “The real business of case study is particularization, not
generalization” (p. 8). In other words, a case study approach entails looking at a particular case and knowing it well.

On the other hand, some researchers provide strategies that are utilized in this study to help with issues of generalizability. First, Merriam (1998) stressed that studies should provide rich, thick descriptions “so that the reader will be able to determine how closely their situation match the research situation, and hence, whether findings can be transferred” (p. 211). Schofield (1990) also provided other strategies to increase qualitative study’s ability for generalization. One technique that is used by this research paper is to study the typical and common sites with the additional value of multi-site studies. This multisite design incorporates two Michigan schools that are demographically representative of the majority of schools in America so that the results from the study can be applied by the reader to a greater range of other situations. Schofield’s other suggestion is to design studies that fit with future trends and issues that are more likely to become typical and common. Teacher evaluation is certainly a movement that is gaining momentum, especially with ongoing federal and state mandates for accountability.

But in the end, Lincoln and Guba (1979) noted, “The trouble with generalizations is that they don’t apply to particulars” (p. 27). Even though I will provide rich and thick descriptive of typical schools in Michigan, qualitative studies, and all research studies in truth by their nature cannot be replicated since the real world changes (Creswell, 2007). Despite these limitations, this qualitative case study design will nonetheless present detailed descriptions of the processes involved in the implementation of a multidimensional teacher evaluation models in two Michigan schools.
CHAPTER IV

FINDINGS

The major concerns about traditional teacher evaluation are its failure to adequately assess the complex and comprehensive scope of the teacher’s position and its failure to have a positive impact on professional growth and leadership practices. The purpose of this study is to investigate teacher evaluation as a promising practice to improve teacher quality and therefore student achievement. This study will more specifically describe how two Michigan schools that utilize Charlotte Danielson’s research-supported Framework for Teaching are attempting to build a multidimensional teacher evaluation system. Using a case study approach, data was collected through interviews, observations, surveys, and document analysis of the schools’ evaluation process. Site visits to the two schools lasted for three days each. The first case school was visited in March 2012 and the second case school was visited in May 2012.

Chapter four is split into four sections. The first section provides an overview of the data analysis process and procedures explained for the document analysis, interviews, and observations. The second section then provides an analysis of the online survey from the two schools. Lastly, the final two sections make use of the data collection to lay out a detailed finding for each of the two case schools. Ultimately, the objective of these case studies is to produce a better understanding of how these two schools that are using research-supported teacher evaluation frameworks are moving toward developing multidimensional evaluation practices and systems.
Overview of Findings

The data collection instruments of document analysis, interviews, observations, and surveys utilize the inductive approach, with the four research questions serving as the guide and protocol. In other words, the data collection instruments are analyzed using the research questions. Table 3 identifies which research question each of these instruments address. For the first research question on how schools are addressing Danielson’s four domains of teacher practice, the following data collection instruments were used: documents, principal, lead teacher, and central office interviews, and classroom and training observations. An understanding of the second research question about resources needed for implementation was gathered through the archival documents, an online survey, and interviews with principal, lead teacher, and central office representative, as well as through the focus groups and classroom observation and training sessions. The interviews, focus group, training observation, and online survey revealed evidences for the third research question about challenges schools faced in implementing the evaluation system. Finally, for the fourth research question, data from the interviews, focus groups, and online survey provided evidences of how the evaluation system has resulted in positive educational outcomes. It is important to note that these triangulations of collected data provide a means to increase reliability and validity by utilizing multiple sources of data (Merriam, 1998; Yin, 2003b).
### Table 3

**Data Sources and Research Questions**

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Research Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Analysis</td>
<td>X</td>
</tr>
<tr>
<td>Principal Interview</td>
<td>X</td>
</tr>
<tr>
<td>Lead Teacher Interview</td>
<td>X</td>
</tr>
<tr>
<td>Central Office Interview</td>
<td>X</td>
</tr>
<tr>
<td>Focus Groups</td>
<td></td>
</tr>
<tr>
<td>Classroom Observations</td>
<td>X</td>
</tr>
<tr>
<td>Training Observation</td>
<td>X</td>
</tr>
<tr>
<td>Online Survey</td>
<td>X</td>
</tr>
</tbody>
</table>

**Data Analysis Processes and Procedures**

As data are collected through document analysis, interviews, and observations, it is important for me to continually reflect about the data and write memos to myself throughout the study. In other words, “qualitative data analysis is conducted concurrently with gathering data, making interpretations, and writing reports” (Creswell, 2009, p. 184). From these ongoing reflections and various methods of data collection, I then attempted to organize and make meaning of the data.
This study approached data analysis with a focus on real-life data. Marshall and Rossman (2011) define this approach as in vivo codes, which are codes that come from the actual data. In other words, this type of data analysis involves segments of text that the researcher lifts out while also trying to stay as close to the data as possible. Using this type of analysis, I organized and prepared data for analysis by transcribing my various interviews, typing field notes from my observations, and optically scanning documents and taking notes. Once data was collected and organized, the data analysis portion was split into three phases. These phases are based upon the overarching approach of inductive analysis using the study’s research questions. The first phase focused on interview and focus group data through the use of inductive analysis of the interview data followed by an inductive analysis of focus group data. The second phase then focused on observation and document data through the use of inductive analysis of the observation data followed by an inductive analysis of the document data. Lastly, phase three focused on data from the online survey through use of descriptive analysis of the survey data. These three phases are listed in Table 4.

Table 4

Data Analysis Plan

<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Inductive Analysis of Interview Data</td>
<td>A. Inductive Analysis of Observation Data</td>
<td>Descriptive Analysis of Survey Data</td>
</tr>
<tr>
<td>B. Inductive Analysis of Focus Group Data</td>
<td>B. Inductive Analysis of Document Data</td>
<td></td>
</tr>
</tbody>
</table>

131
Each step of the above data analysis plan involved reading through the appropriate data several times and then analyzing it through a coding process. “Coding is the process of organizing the material into chunks or segments of text before bringing meaning to information” (Creswell, 2009, p. 186). These segments of text came from clustering similar topics and then abbreviating the topics as codes. Stake (1995) viewed this step as searching for meaning by searching for patterns and consistencies. It should be noted that Marshall and Rossman (2011) reminded the researcher, “this phase of data analysis is the most difficult, complex, ambiguous, creative, and fun” (p. 211). With considerable time devoted to immersing myself into the data, categories that are alphabetized as codes eventually materialized. As with in vivo coding, the categories that are marked come from meaning units or actual phases used in specific text segments. It is important to note that these codes will emerge during the data analysis process and not from predetermined codes. In fact, some of these codes and categories reinforced past literature while others were not anticipated at the beginning of the field study. After identifying categories, the next step involved continually revising and refining the category system so that contradictory points of view and new insights are included with appropriate quotes that express the central theme or essence of a category. Categories were also combined when meanings were similar. Creswell (2002) provided the chart in Figure 2 to represent the aforementioned coding process for inductive analysis.
Figure 2. Coding process for inductive analysis.

From the coding process in phase one and two, the next section of this chapter begins with a detailed description of the setting and people, which are vital for case study approaches (Creswell, 2009). These descriptions are meant to give the reader a sense of being present at that setting and time. As Stake (1995) pointed out, “The physical space is fundamental to meanings for most researchers and most readers” (p. 63). I identified common themes and categories through the three phases of the data analysis plan, which are subsequently used as headings in the following sections of this chapter. Since all the data has been coded and grouped together, these themes and headings included multiple perspectives with quotations and evidence from interviews, focus groups, observations, document analysis, and surveys. Furthermore, because this study includes case studies at two different sites, themes are organized according to each individual case. When possible, I will make connections between themes across the two sites in the next chapter. In other words, themes will be represented in this study through detailed discussion of the major themes and subthemes at each site with the addition of interconnecting themes. Merriam (1998) identified this approach when using a multiple case study, where “there are two stages of analysis—the within-case analysis and the cross-case analysis” (p. 194). Figures and tables will supplement the findings as well, such as visually showing
differences and similarities in themes across the two sites and how the data sources correspond to the research questions.

**Overview of Survey Results**

Before examining findings for each school, this section will provide an overview of the online survey. The online survey (Appendix F) was administered through Google using 4-point Likert-type scale questions of strongly agree (4) to strongly disagree (1) with the option of “don’t know.” Questions were focused on exploring teacher perceptions regarding the use of their teacher evaluation system. Specifically, the survey items were based upon the four essential attributes of sound and fair evaluations, according to the Joint Committee on Standards for Educational Evaluation (2009). These attributes include:

1. Propriety standards to ensure legal, ethical, and due regard for the welfare of the evaluatee and those involved in the evaluation.
2. Utility standards to guide evaluations so that they will be informative, timely, and influential.
3. Feasibility standards to ensure that the system is as easy to implement as possible, efficient in their use of time and resources, adequately funded, and viable from a political standpoint.
4. Accuracy standards to determine whether the evaluation has produced sound information and are technically adequate and as complete as possible (p. 6-7).

Essentially, these standards are utilized to examine the efficacy of the teacher evaluation model at each school. In addition to these items, general questions were included regarding professional development and teacher’s perception of the school’s previous
evaluation model. The survey’s final section included a free response area for personal responses on the major advantages and disadvantages of the evaluation process.

In addition to the survey items coming from the above-mentioned four attributes, the survey statements also connect with the four research questions from this study. For instance, statements under the fairness, usefulness, and accuracy categories of survey tie into the first research question, “How are schools addressing each of Danielson’s four domains of teacher practice as they begin to make their evaluation practices and systems more multidimensional and transformational (growth-oriented)?” The last two statements listed under the feasibility category of the survey regarding the written guidelines, procedures, and resources of the evaluation system connect with the second research question, “What resources are utilized to implement a multidimensional teacher evaluation system?” The other three statements listed under the feasibility category about the amount of time and training tie into the third research question, “What challenges do schools face in implementing a multidimensional teacher evaluation system and how are they addressed?” Lastly, the survey’s general questions portion regarding professional growth and changes in student achievement connect with the final research question, “What evidence(s) exists that the teacher evaluation system has or is resulting in positive educational outcomes?”

The online survey link was emailed to the staff at each school via a group email distribution list, with participation being voluntary and anonymous with no benefit provided for those who chose to participate. Teachers at both campuses had 10 days to complete the survey. The following sections that examine the two case schools will also include a portion tied to the online survey results.
Research Findings: Two Case Schools

The final sections of this chapter will examine research findings for each of the two case schools, which are connected to the four research questions. Each portion begins with an overview and background of the school, followed by a description of their evaluation process. Such a description will include history and goal of the school’s evaluation system and implementation details. Then the school’s online survey results will be examined. The final section for each school will attempt to answer the four research question related to the evaluation system’s use of the Danielson framework, resources, challenges, outcome, and lessons learned.

Case School #1

Background of the School

With a history of over 100 years, the high school is located in a semi-rural neighborhood with close access to nearby, larger suburban communities. The school’s mission is as follows: “In an environment of life-long learning, we work together to recognize and develop potential and to promote success” (School website). In order to achieve its mission, the school is committed to providing a safe and stimulating learning environment delivered through a comprehensive curriculum with high student expectations from staff members who are competent, caring, and enthusiastic.

The high school has an enrollment of 695 students from ninth to twelfth grade. The district has one of the highest household incomes in their county since many of its residents are in the entrepreneurial and professional line of work (School website). The majority of students are white with about 12% eligible for free or reduced lunch (School website). The school follows a trimester schedule with five class periods of seventy-three
minutes. Staff includes 34 teachers, 2 administrators, and 3 para-professionals. The principal has held the position since 2004 and was also a former county teacher of the year. He shares operational and instructional responsibilities with his assistant principal, who has served in the position since 2002. Detailed demographics of the school are provided in Table 5 below.

Table 5

*Demographic Profile of High School (2011-2012)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grades served</td>
<td>9&lt;sup&gt;th&lt;/sup&gt; to 12&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Students served</td>
<td>695</td>
</tr>
<tr>
<td>Student population breakdown</td>
<td>White, not Hispanic: 99%</td>
</tr>
<tr>
<td></td>
<td>Hispanic: &lt; 1%</td>
</tr>
<tr>
<td></td>
<td>Asian/Pacific Islander: &lt; 1%</td>
</tr>
<tr>
<td></td>
<td>Black, not Hispanic: &lt; 1%</td>
</tr>
<tr>
<td>Students eligible for free or reduced-price lunch program</td>
<td>12%</td>
</tr>
<tr>
<td>Number of full-time administrators</td>
<td>2</td>
</tr>
<tr>
<td>Number of teachers</td>
<td>34</td>
</tr>
<tr>
<td>Per-pupil spending</td>
<td>$7130</td>
</tr>
</tbody>
</table>

**Goal of the Teacher Evaluation System**

The main goal of the teacher evaluation system is to improve teacher effectiveness and teacher reflection so student achievement improves. It is based upon the principle of “what would make weak teachers strong and good teachers great” (Participant A, principal). Such improvement comes through a system set up to provide greater opportunities for self-reflection and collaboration. In other words, the system is
committed to “developing an evaluation process that supports the continual improvement of our instructional practices and collaboration of all stakeholders” (School website). It is important to note that the central office representative emphasized to his staff from the first day that this system cannot be punitive but instead is focused on professional development (Participant B). The intention is for teachers to reflect and learn through multiple opportunities provided by instruments such as peer coaching and portfolios combined with administrative support as avenues towards becoming better teachers.

**Implementation of the Teacher Evaluation System**

**History.** The evaluation system has been in place at the high school for two years. Beforehand, the previous system was based on a single sheet evaluation form with three levels of rating: high, medium, or low (Participant B, central office). Principals used to observe teachers two times a year with a checklist of items to spot, such as bulletin boards in the classroom. Information from these formal observations determined teachers’ final evaluation, which was presented to them and signed off at the end of the year. According to the principal, central office representative, and lead teacher, the former system lacked accountability (Participant A; Participant B; Participant C). For instance, the principal mentioned that tenured teachers especially were not motivated by the system to make changes because it would cost the district too much money to even consider marking them down (Participant A).

The major catalyst for beginning the process of changing the previous system was “rumblings from the state” (Participant A, principal) that a new evaluation model may be forced upon districts if they do not develop their own system. As one teacher in the focus group said, the consensus among staff members at the time they heard that Michigan may
implement a new teacher evaluation system was to say: “let's determine our own fate; let's write our own contract” (Participant F).

Soon thereafter, the superintendent formed an evaluation system committee consisting of a teacher from each campus, the teacher’s union president, the upper elementary and high school principal, and the superintendent. They met every other Monday for two hours. A major portion of that time was spent on determining what framework to use. After examining various options, the committee decided to use the Danielson framework for several reasons. First, it provided a common standard that all campuses in the district could use (Participant A, principal). Second, the committee felt that the framework was also comprehensive, with the domains representing areas of teaching that could be evaluated (Participant C, lead teacher; Participant B, central office). Additionally, the Danielson domains were neither overly technical nor conceptual but were instead explicit on what each area of teaching looks like so that teachers specifically know expectations (Participant C, lead teacher; Participant A, principal). The principal contrasted it from to the old system, which was similar to the Potter Stewart definition of pornography that could only be defined when he saw it (Participant A). On the contrary, the Danielson framework “spells out” the various aspects of teaching so that teachers know what to do (Participant A).

After one year of the committee meetings, the new system was developed. It was then presented to the staff by the union president with feedback from teachers further considered by the evaluation committee (Archive A). Forty volunteer teachers were selected to take part in the first year as a pilot. After the first pilot year and further modifications to the system, the high school implemented it school-wide in fall of 2011.
Description of teacher evaluation system. The high school’s teacher evaluation system begins with a group informational meeting at the beginning of the year for all staff members. The principal conducts this meeting for those interested in learning more about the evaluation process. Since returning teachers are typically aware of the process, the meeting is optional. During this meeting, the principal goes over the Danielson framework and all components of the evaluation system with an opportunity to answer any staff questions or concerns.

The structure of the evaluation system is based upon two separate tracks for staff members depending upon if they are on their focus year. For a beginning teacher up to the point of receiving tenure, and every three years thereafter, teachers are considered to be on their focus year for evaluation. In the focus year, teachers must be evaluated by an administrator observing their classroom through the use of the Danielson framework. These evaluations must be completed a minimum of twice yearly with a pre- and post-conference (Archive A).

In the pre-conference, teachers meet with their principal during their planning period for 30 minutes. The purpose of this meeting is to allow teachers to provide an overview and the goal of the lesson to be observed by the principal (Archive B). The teacher is required to bring a lesson plan outline to guide this conversation. The principal may also bring up areas on concern from previous observations so that the teacher is made aware of possible focal points in the upcoming classroom observation.

During the classroom observation, which takes place for a full class period, the principal uses the Danielson framework to guide his observation with comments and notes entered in the appropriate domain. The post-conference is then scheduled within
three days of the classroom observation. The principal structures this meeting by going through the different parts of the lesson observed with feedback provided from the observation form (Archive C). The teacher has opportunities to provide their comments if they so desire. This meeting ends with the final evaluation score shared with the teacher.

Besides the above mentioned classroom observations during focus years, all teacher are also required to submit a portfolio by April 15th every three years. The portfolio component will be explained later.

For a tenured teacher who is not in their third-year cycle of teaching, they are categorized in the second evaluation track known as the non-focus year. These teachers are observed by the administrator with classroom walkthroughs, and then they have the option of completing one of four evaluation instruments: peer coaching, lesson study, video lesson, or action research. The only stipulation is that a teacher cannot choose the same instrument more than two years in a row (Participant A, principal).

Figure 3 shows a flowchart of these two evaluation tracks for all staff members in the high school. Each component of the below evaluation system flowchart is described individually thereafter.
Figure 3. Flowchart of two tracks of high school evaluation system.

**Portfolio.** The portfolio is a collection of artifacts that demonstrate the work of the teacher so that they are able to show what they are actually doing in the classroom (Participant A, principal). It must be submitted every three years, affording teachers time to collect artifacts. A list of these potential artifacts, which are connected to the four domains of the Danielson framework, is included in the Portfolio Artifact Matrix (Appendix E). This matrix essentially serves as an organizational tool for teachers to record all artifacts they collect.

On the other hand, the evaluation manual and the lead teacher who was interviewed emphasized that the format for the portfolio is also “fluid and dynamic” to reflect each teacher’s practice (School website; Participant C). In other words, a standardized portfolio format is intentionally not provided because it should be
determined by each teacher based upon their specific situation. The only requirement is that they include artifacts from all four domains of the Danielson framework. The deadline to submit their portfolio is by April 15\textsuperscript{th} every three years.

**Peer coaching.** In order to build collaboration and improve teacher practice, peer coaching is one of the four optional instruments that teachers can complete during their non-focus years. The teacher is provided the opportunity to select a professional colleague for peer coaching, as long as they first obtain approval of the selection from their administrator. Furthermore, the coaching partner must not be a colleague that the teacher had already partnered within the last two years. In essence, this instrument makes use of teachers coaching each other. The peer teachers will schedule two dates for peer coaching to occur, once before winter break and the other before April 15\textsuperscript{th}. The school allows teachers to obtain substitutes during those days as long as they make the request to their supervisor at least five days prior.

Before each coaching session, teachers must prepare by conducting a planning meeting with their partner. In this pre-meeting, the teacher will provide an outline of areas for their coaching partner to focus on and review the lesson. Then the coaching visit will last for a minimum of 45 minutes. Coaches have the opportunity to use the Coaching Observation Form (Appendix G) based upon the Danielson framework. Thereafter, the teacher will review notes and suggestions for improvements with their peer colleague during a post-coaching debriefing. And finally, the teacher will complete a Professional Reflection Sheet (Appendix H) and submit this document by winter break and April 15\textsuperscript{th}.
It should be noted that comments from the coach are optional and not required to be submitted. Additionally, peer coaching is meant to be non-evaluative and instead focused on helping teachers meet their professional goals. For instance, teachers in the first focus group shared an example of the administrator informing the peer coach of a teaching behavior that was of concern so that the coach can help their teacher improve in that specific area (Participant F, teacher). Consequently, the peer coaching component has an indirect affect on the evaluation since it provides another tool to improve teachers’ professional practice while also assisting administrators.

**Video coaching.** As another opportunity for teacher self-reflection, the second optional instrument of choice during non-focus years is video coaching. Teachers first determine which lesson they will record. Then they reserve a web camera that is designed to record the entire space of a classroom (Participant A, principal). Once the lesson is recorded, the teacher will then review it using the same Danielson framework that administrators use during their classroom observations. Lastly, the teacher will submit the Video Reflection Form (Appendix I) to their administrator. These steps must be completed at least two times during the year, with the same deadline of Winter break and April 15 for submission. The video lesson may also be archived by the teacher for later reference.

**Lesson study.** As another opportunity for teacher collaboration and reflection, the third option is a lesson study. According to Stepanek, Appel, Leong, Mangan, and Mitchell (2007), a lesson study “is a professional development practice in which teachers collaborate to develop a lesson plan, teach and observe the lesson to collect data on student learning, and use their observations to refine their lesson” (p. 2). Teachers can
select a team of 2 to 6 staff members. These teachers will use the above mentioned book by Stepanek and colleagues as a guide to conducting their lesson study.

Once a schedule is laid out with times to meet, the team will begin their first gathering by determining the focus and topic of the lesson. This determination should take into consideration areas of weakness on local and/or state level assessments in the subject area (Archive A). The topic must be determined by October 1. Then once the lesson has been developed, it will be recorded on the Lesson Study Template (Appendix J). It is also suggested that the team take notes from each meeting and include it at the beginning of the template. This template document is the first submission due to administrators before Winter break.

After the lesson is developed, the team determines when and who will present it. Teachers can seek administrator approval for substitutes provided that it is requested a minimum of five days prior. Teachers then utilize the Danielson framework while observing the lesson. Thereafter, the team conducts a debriefing, with the goal of adding revisions to the original template document. If desired, the teachers can present a second time to see the effectiveness of the revisions. The updated and final lesson template will be submitted by April 15.

**Action research.** The final instrument of choice for further collaboration and self-reflection of teacher practice is action research. Similar to lesson study, this instrument is completed as a team with two to four teachers. The goal of action research is to understand some element of the teachers’ classroom by collecting data. It is based upon the assumption that “teachers and principals work best on problems they have identified for themselves” (School website). Data collection should be based upon
multiple incidents taken at different times and in a variety of ways. Examples of these data collections include a log or research journal, interviews, individual or small group student conferences, and quantifiable data such as assessment scores. Data is first collected in order to identify an area of focus. Then the team interprets that data and forms an action plan.

For the high school evaluation system, the teachers will first decide what topic they want to study by October 1. The topic can be modified as the process continues. Then by the first of November, they will know what data they want to collect and then start collecting it while also keeping notes throughout the process. By December 14, teachers will report status of progress to their administrator. All data must be collected and data analysis will begin by March 15 with a written report generated by April 1. The final deadline is April 15, whereupon the team will submit their final product to their administrator as they also prepare materials for a staff-wide presentation. The date of the staff presentation will be determined by administrator.

In some cases, the action research project can last more than one year, if approved by the administrator, with a maximum of two years allowable. Such a plan should be outlined briefly for the administrator at the onset of the project. And along the way, brief reports should be provided before Winter break and April 15.

Final evaluation. The final evaluation for all teachers will take place by May 15th. This session is led by the principal, who will go through the teacher evaluation worksheet to outline scores in each area of the evaluation system. The overall performance range is divided into four categories: highly effective, for 90-100%,
effective, for 75-89%, minimally effective, for 60-74%, and not effective, for 0-59%.

These percentages are based upon the following two sections of the evaluation.

**Danielson’s domains and evaluation instruments.** The first portion of the evaluation system accounts for 50% of the total evaluation score. This percentage is based upon the four domains of the Danielson framework and the chosen evaluation instrument for that year. Specifically, each of the four Danielson domains will count for 10% for a total of 40%. The last 10% will come from the chosen evaluation instrument, based upon the number of years teaching, consisting of peer coaching, lesson study, video lesson, action research, or portfolio. Table 6 shows a breakdown of these percentages.

Table 6

**Danielson’s Domains and Evaluation Instruments**

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1: Lesson Planning</td>
<td>10%</td>
</tr>
<tr>
<td>Domain 2: Classroom Environment</td>
<td>10%</td>
</tr>
<tr>
<td>Domain 3: Instruction</td>
<td>10%</td>
</tr>
<tr>
<td>Domain 4: Professional Responsibility</td>
<td>10%</td>
</tr>
<tr>
<td>Chosen Evaluation Instrument</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Data component.** In addition to the above instruments that teachers must complete either during their focus or non-focus year, the high school evaluation system also includes a data component that is completed by every teacher, every year. This component accounts for the other 50% of the overall evaluation score. It is based upon growth percentage from the beginning of year to the end of year from national, state, and
local assessments. Thirty percent of the data component will come from NWEA’s national Measures of Academic Progress (MAP) assessment. Five percent will be based upon state tests such as MEAP or MME, and fifteen percent come from local assessments created by the district. A breakdown of these percentages is shown in Table 7.

Table 7

**High School Evaluation System: Data Component**

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Assessment (NWEA)</td>
<td>30%</td>
</tr>
<tr>
<td>State Assessment (MEAP or MME)</td>
<td>5%</td>
</tr>
<tr>
<td>Local Assessment (District-Generated)</td>
<td>15%</td>
</tr>
</tbody>
</table>

**Survey Results: Case School #1**

For the first school’s online survey, 24 teachers submitted responses out of 34 teachers from the high school campus, for a response rate of 70.6%. A majority of experts in the field recommends a 60% response rate for generalizability of results (Glatthorn, 1998), which was achieved among teachers from this first case study.

The actual response numbers and percentages are shown in Tables 8-10. First, Table 8 provides a breakdown of the initial two questions in the survey where teachers were asked to indicate amount of classroom teaching experience at current campus and also total number of years of teaching experience. This table shows the number of teachers considered new teachers (those with five years of experience or less) compared to veteran teachers. A mix of both new and veteran teachers are evident from survey results for the high school.
Table 8

Distribution of New and Veteran Teachers: Case School #1

<table>
<thead>
<tr>
<th></th>
<th>Number of Teachers with Years of Teaching at Current Campus</th>
<th>Number of Teachers with Total Years of Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Teacher</td>
<td>10 (42%)</td>
<td>10 (42%)</td>
</tr>
<tr>
<td>Veteran Teacher</td>
<td>14 (58%)</td>
<td>14 (58%)</td>
</tr>
</tbody>
</table>

The next portion of the survey included 19 Likert graded statements. As previously mentioned, these statements were based upon the four essential attributes of sound and fair evaluations, according to the Joint Committee on Standards for Educational Evaluation and connect with specific research questions. Tables 9-12 display the percentages of the 24 teachers who provided responses in the Likert scaled section of the survey, with statements that are condensed from the original survey and categorized according to the four research questions. Table 10 shows respondent’s Likert scale percentages for statements under fairness, usefulness, and accuracy categories that relate to the first research question about the makeup of the evaluation system that is based upon the Danielson framework. Overall teachers provided positive feedback about the usefulness and accuracy of model, especially with regards to the Danielson framework. On the other hand, there was less agreement from teachers about the overall fairness and efficiency of the evaluation system measuring their level of performance.
Table 9

Statements Categorized Under School Addressing Danielson’s Domains (Research Question #1)

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair and Efficient</td>
<td>8%</td>
<td>46%</td>
<td>42%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>Prominent Role</td>
<td>17%</td>
<td>63%</td>
<td>21%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Two-Way Communication</td>
<td>13%</td>
<td>58%</td>
<td>21%</td>
<td>0%</td>
<td>8%</td>
</tr>
<tr>
<td>Promotes Good Teaching</td>
<td>21%</td>
<td>50%</td>
<td>25%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>Encourages Self Reflection</td>
<td>33%</td>
<td>67%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Change Instructional Practices</td>
<td>17%</td>
<td>58%</td>
<td>21%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>Identify Strengths and Weaknesses</td>
<td>13%</td>
<td>71%</td>
<td>13%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>Danielson Provide Accurate Representation</td>
<td>13%</td>
<td>71%</td>
<td>13%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>Provide Evidence</td>
<td>13%</td>
<td>63%</td>
<td>21%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>Richer and Comprehensive</td>
<td>13%</td>
<td>58%</td>
<td>25%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>Adequate Data Sources and Inputs</td>
<td>8%</td>
<td>50%</td>
<td>21%</td>
<td>4%</td>
<td>17%</td>
</tr>
</tbody>
</table>

Table 10 shows respondent’s Likert scale percentages for two statements under feasibility category that connect with the second research question, “What resources are utilized to implement a multidimensional teacher evaluation system?” Teachers
generally submitted a lesser degree of agreement about clarity and availability of written policies with about 33% agreeing, and most teachers, with 67%, either disagreed or strongly disagreeing.

Table 10

*Statements Categorized Under Use of Resources (Research Question #2)*

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearly Written &amp; Provided Policies</td>
<td>4%</td>
<td>29%</td>
<td>50%</td>
<td>17%</td>
<td>0%</td>
</tr>
<tr>
<td>Sufficient Resources</td>
<td>0%</td>
<td>54%</td>
<td>21%</td>
<td>13%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Table 11 displays respondent’s Likert scale percentages for the first three statements under the feasibility category that connect with the third research question, “What challenges do schools face in implementing a multidimensional teacher evaluation system and how are they addressed?” Though half of the respondents at least agreed that there is reasonable time to review evaluations and that their evaluator was qualified, most of these teachers also expressed concerns about the insufficient training provided by the district for teachers.

Table 11

*Statements Categorized Under Challenges Encountered (Research Question #3)*

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Reasonable</td>
<td>4%</td>
<td>54%</td>
<td>17%</td>
<td>8%</td>
<td>17%</td>
</tr>
<tr>
<td>Teacher Training</td>
<td>0%</td>
<td>25%</td>
<td>58%</td>
<td>17%</td>
<td>0%</td>
</tr>
<tr>
<td>Sufficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluators Trained</td>
<td>8%</td>
<td>38%</td>
<td>25%</td>
<td>13%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lastly, Table 12 displays respondent’s Likert scale percentages for the final three statements of the survey that connects with the final research question, “What evidence exists that the teacher evaluation system has or is resulting in positive educational outcomes?” Though a far majority of teachers disagreed that the former model led to professional growth, these teachers also provided mixed feedback or uncertainty around the new model’s impact on their professional growth. Nonetheless, 58% of teachers expressed agreement that the new evaluation system will lead to student growth.

Table 12

*Statements Categorized Under Evidence of Growth (Research Question #4)*

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Former Model Provide Growth</td>
<td>0%</td>
<td>13%</td>
<td>46%</td>
<td>29%</td>
<td>13%</td>
</tr>
<tr>
<td>New Model Lead to Student Growth</td>
<td>33%</td>
<td>25%</td>
<td>21%</td>
<td>8%</td>
<td>13%</td>
</tr>
<tr>
<td>New Model Provide Appropriate Training</td>
<td>4%</td>
<td>42%</td>
<td>4%</td>
<td>17%</td>
<td>33%</td>
</tr>
</tbody>
</table>

The final portion of the survey included an optional free response section for respondents to provide their personal views regarding major advantages and disadvantages of the evaluation process. From the 24 teachers who submitted an online survey, 12 teachers (50%) submitted their thoughts on the major advantages of the evaluation system. Additionally 14 teachers (58%) also provided their views on the major disadvantages of the evaluation system. In order to evaluate the free response section, each submission was categorized from the transcripts for word frequency and thematic phrasing.
After such analysis, three major themes were present in the first section that relates to the major advantages of the evaluation process. Six teachers made reference to opportunities for collaboration with statements such as “it promotes interaction with administrators and teachers” and “it also provides opportunities to work more closely with colleagues”. The second most common theme on advantages, with four teachers, was that the evaluation system reflects teacher practice. Statements for this theme include “Let’s administrators know some of things that a teacher has done” and “provides some sort of documentation of my performance”. Three teachers referred to the last common theme of promoting reflection with statement such as “The evaluation process provides for many opportunities of self-reflection” and it “enables one to reflect on a professional organizational system”.

Three major themes also came from the analysis of the second section that relates to the major disadvantages of the evaluation process. Seven teachers referred the issue of time consumption with comments such as “this process is one more thing on our plates when so much is being heaped on already” and “There is not enough time given to complete the evaluation process thoroughly”. Four teachers questioned the legitimacy of the data component with statements such as “I don’t think there is enough consideration of the students responsibility incorporated into the evaluation” and “Many students do not take the NWEA testing as seriously as they should hence they don't put forth their best effort; there are no consequences for refusing to try their best leading to possible skewed results.” The final most common theme mentioned by three teachers related to the evaluation system being complicated with a comment such as “Because it is a new process, I wasn't sure about some of the artifacts and I had to ask a lot of questions.”
Overall, the free response section provided a voice to the simple markings from the Likert scale statements of the survey.

**Research Questions Examined**

**Question #1: Danielson framework.** As a review of the first research question, the Danielson framework “identifies those aspects of a teacher’s responsibilities that have been documented through empirical studies and theoretical research as promoting improved student learning” (Danielson, 1996, p. 1). It is organized into four domains of activity that refer to distinct aspects of teaching. These domains provide a common definition of expertise that the school can utilize when evaluating teacher practice. In fact, all individuals interviewed for case school #1 responded that the Danielson framework was chosen because it offered a common research-proven standard of what effective teaching looks like (Participant A, principal; Participant B, superintendent; Participant C, lead teacher). Furthermore, 84% of surveyed teachers at this school also agreed that the Danielson model provides an accurate representation of teaching.

In the high school evaluation system, the first three Danielson domains of planning and preparing, the classroom environment, and instruction are covered during the principal’s classroom observation and also the optional evaluation instruments that teacher submit during their non-focus years. For classroom observations, the principal ensures that all their comments and feedback are based upon the first three domains. In fact, the observation form is split into the three domains with respective components listed. The first domain of planning and preparing is completed by the principal after looking at the lesson plan and considering the comments made by the teacher during the
pre-conference. Then during classroom observations, the principal notes his reflections through the lens of the other two domains of classroom environment and instruction.

Additionally, the chosen evaluation instrument during the non-focus years is built around the first three Danielson domains. Peer coaching utilizes the Coaching Observation Form (Appendix G), which is similar to the tool that the principal uses when observing lessons. In doing so, teachers are provided an opportunity to observe each other in a similar fashion as their administrators. Secondly, the video coaching provides another chance for teachers to self-reflect on their own practice by reviewing their recorded lesson through the lens of the first three Danielson domains. The Video Reflection Form (Appendix I) that teachers submit is built around the Danielson rubric as teachers observe their behavior according to the various components of the framework. For instance, the first question in the form asks teachers to examine and reflect upon each component of the classroom environment and instruction.

The third evaluation instrument of lesson study specifically focuses, particularly at the onset, on the first domain of planning and preparing. It is here that teachers work together to develop a lesson plan while at the same time reflecting upon the Danielson components under the first domain. In fact, the Lesson Study Template (Appendix J) is built around the same domain. For instance, the unit objective and outcome in the lesson plan template connect with the first and third components of the primary domain. The question regarding current characteristics of students tie into the second component while the question concerning the sequencing of lessons connects with the fifth component. Finally, questions about lesson plan materials and data collection tie into the forth component while the question about the learning process relates to the last component.
After planning the lesson, teachers make use of the second and third domains of Danielson framework as the lesson is taught and observed, and soon thereafter when it is refined.

The last evaluation instrument of action research ties into any one of the Danielson domains depending upon which problem the teacher will examine, which can range from issues with planning, the classroom environment, or instruction.

The final Danielson domain of professional responsibilities is primarily handled through the portfolio that teachers put together and submit. This instrument will allow teachers to include artifacts that connect with the six components of the forth domain. For instance, the portfolio must include accurate records and communications with family while also showing evidence of participating in and improving professional practice. By collecting and putting together their portfolio on an ongoing basis, teachers are also reflecting upon their practice of professional responsibilities along the way as well.

**Question #2: Resources.** By examining the various resources utilized in the high school evaluation model, though it is difficult to calculate a percentage of time spent by staff members on the evaluation process, it is nonetheless evident that time is the first major resource. Initially, the evaluation committee met for a year every other Monday to develop and form the evaluation model. Serving as volunteers, there were not any monetary costs involved for the district. Then during the evaluation process, the principal was the main staff member who spent about 30% of his time in conducting the evaluation, which is equivalent to 2 hours a day (Participant A, principal). Of course, in addition to implementing the evaluation system, he was also responsible for the operation and instructional programs of the school, with the help of his assistant principal. Another
prominent staff member was the lead teacher who represented the high school campus in the evaluation committee. She served as the resident expert whom the teachers can ask questions and express concerns. Additionally, there were 40 pilot teachers who played a role in further developing the evaluation model during the first pilot year. Both the lead teacher and pilot teachers served as volunteers. The other staff members who took part in the evaluation process found time during and outside school hours to complete the various components that were required of them. This includes the coaches, who were also full-time teachers.

Besides the resource of time, there were a few direct costs that the district had to bear to implement the evaluation system. One budgetary consideration was the cost of substitutes for teachers who requested their help to complete various evaluation instruments such as peer coaching or any of the collaborative instruments. But high school teachers in the focus group mentioned that they did not request substitutes since they either used their planning period or their time before or after school (Participant F). Nonetheless, budgetary considerations for substitute teachers were incorporated as a necessary resource for the district. Another major budgetary cost for the district is the time and money spent on training focused on the evaluation system. Initially, consultants were brought into the district to train administration and staff. These consultants were used for one day of training per school year. For instance, consultants from NWEA trained the staff members about the MAP tests that would be used as the national assessment (Participant B, central office). In this respect, consultants had an indirect role in the evaluation process by assisting in its implementation. Furthermore, the district offered training for all staff members on the evaluation process for two days at the
beginning of the year. Then each component of the evaluation process was covered in the first five Monday’s of the school year (Participant A, principal). Additional training was provided by each campus, as needed, throughout the year.

Besides time spent for these various trainings, the district also purchased a few resources to implement the evaluation system. For instance, the district pays a fee every year to use the online MAP assessment, which serves as the pre- and post-national assessment in the data component of the evaluation system. Secondly, the district compiles and analyzes data by using Stages with a data component called Informed. Additionally, five cameras were purchased so teachers can complete their video recordings. It should also be mentioned that these resources require computers for administrators and teachers to complete their evaluation steps. Students also needed computers to complete the online MAP assessment and any other online assessments. But these resources are not additional costs for the district since administrators and teachers are already provided laptops and students have access to desktops in the high school computer lab.

**Question #3: Challenges.** There were a few challenges that came about from the high school teacher evaluation model. The central office representative, principal, and teachers shared some of their personal and professional challenges with the teacher evaluation system and how they attempted to resolve it, if possible. These challenges included time constraints, teacher buy-in, and professional development and training.

**Time constraints.** The most prevalent challenge heard from the administrators and teachers was the extra time commitment required of the new evaluation system. It began with the evaluation committee. The principal and central office representative
admitted that they initially planned for the committee to meet for a couple of months to
develop the system (Participant A; Participant B). As it turned out, the committee met
instead for a complete year. And they are still meeting as necessary to improve the
system. So the principal and lead teacher suggested that other districts looking into
developing an evaluation system should plan to spend at least a year or use the summer
breaks to develop the system’s structure.

Once the system was implemented, teachers expressed their concern with finding
time to balance their workload with the new requirements of the evaluation system
(Participant E; Participant F; Participant G; Participant H). They now have to meet with
their principal more often, remember to continually collect artifacts for their portfolio,
and then find time to complete their chosen evaluation instrument. As one teacher wrote
on their survey response, “it is very time-consuming”. A major reason expressed by
teachers for the time consumption is the required paperwork (Participant E; Participant
G). Each evaluation component requires submitting at least one form to their
administrator. On the other hand, as one way to address this challenge, the evaluation
system provides options for teachers and does not require completing multiple evaluation
instruments in the same year. Alternatively, a couple of teachers in the focus group and
online survey labeled this challenge of completing paperwork as a “necessary evil”
(Participant F; Survey). Ideally someone could observe what he or she does every
moment of every day, but since this is not possible, the paperwork is a necessary process
to document teacher practice for the evaluation model.

Teacher buy-in. The second most common challenge cited by administrators and
teachers was staff buy-in of the evaluation system. The principal and lead teacher
admitted that this buy-in was easier for the evaluation committee members since they realized that the state would be mandating their evaluation system if districts did not create one themselves (Participant A; Participant C). So they understood the advantages of developing a system in-house. On the other hand, once the system was developed, teachers were at first skeptical. First, teachers questioned the intent of the new evaluation system. Some feared that it would be used as a tool to remove teachers (Participant F). They were especially worried about the data component. Teachers in the focus group and survey questioned how teachers performance could be tied to test scores (Participant I; Participant F; Survey). They asked how teachers could be accountable for what students do when there are many other factors that affect students’ lives. What about external factors such as students’ biological changes and family and money problems? (Participant F) Though teachers admitted that they play a role in the child’s education, they are not the only factor. As one teacher said, “I don’t know that it’s necessarily scientific” (Participant E). This teacher saw teaching not as a scientific endeavor but rather an art that cannot necessarily be measured.

To address the concern of teacher buy-in, the district attempted to involve teachers as much as possible in the development of the evaluation system. First, a teacher from each campus served in the committee; this teacher had to be an influential staff member who was well-respected and nominated by the other teachers (Participant A, principal). And when the system was developed, the campus completed a pilot year with 40 teachers who volunteered. From the experiences of those teachers, further modifications to the system were made by the committee. Then during its full implementation, the committee made it clear to teachers that they are open to further
suggestions from teachers and that “it’s not set in stone” (Participant F, teacher). The central office representative emphasized as a key lesson learned, “you have to involve the people that you are evaluating in the development of the model. It was an input, input, input to develop that relationship with others” (Participant B).

In addressing challenges regarding the data component of the evaluation system, the central office representative first clarified that teachers have to be held accountable; this accountability “primarily comes down to the academic progress of their students” (Participant B). It is the responsibility of teachers to address the needs of their students so that learning and academic achievement is central to their practice; in his words, the central office representative stated directly that this “just makes common sense” (Participant B). Therefore, the data component is built around the goal of achieving one year of academic growth, which is measured by pre- and post-tests around a mix of national, state, and local assessments (Participant A, principal; Participant C, lead teacher). In addition to using multiple assessments to track this progress, the principal also pointed out that the first year will serve as the base year and therefore data scores will not be a part of summative evaluation. Though teachers will receive their data from assessments the first year, the data component will only be included the following year in the evaluation model (Participant A, principal). Therefore, it is a growth-based model that examines year to year trends. Additionally, it is important to note that the data portion is only 50% of the evaluation model and therefore is not the sole criterion upon which teachers are evaluated.

**Professional development.** The third most cited challenge from administrators and teachers was the need for more professional development related to the evaluation
system. When the model was first implemented school-wide, the high school spent the first two days of the school-year training teachers on the evaluation process. Then for each of the first five Monday’s, trainings were offered after school and focused on one of the five evaluation instruments. Despite these trainings, 75% of teachers nonetheless disagreed on the survey that the professional development and training on the evaluation system for teachers were sufficient. In fact, the principal also agreed that his school did not provide enough training focused on the evaluation system (Participant A). But he also mentioned that there are not enough professional development days in the calendar to offer more training. To add to these complications, the Monday trainings were optional since it was outside the teacher contract.

The insufficiency in professional development was made even more evident during the focus group sessions with teachers. They expressed concerns about what was expected of them. A commonly expressed word was “confusion” especially considering that the evaluation system offered a number of options for teachers in their non-focus years (Participant C; Participant G; Participant H; Survey). They thought that the training at the beginning of the year was inadequate since they were not sure which instrument they would use at that point. Plus, it was a lot of information presented in the first two days. Furthermore, some teachers thought that the open-endedness of what was required, especially in the portfolio, caused further confusion (Participant C; Participant H). They would have like to see examples and models of completed evaluation instruments. These multiple challenges expressed by teachers speak once again to the insufficient training on the evaluation system that has not been addressed so far.
**Question #4: Evidence of impact.** The main goal of the high school teacher evaluation system is to improve teacher practice through feedback on classroom teaching combined with further opportunities for teacher collaboration and self-reflection, which will subsequently show results in student achievement growth. In the words of the principal, the system is built around “trying to develop better teachers, trying to get teachers, even good teachers, to improve their craft” (Participant A). Eighty percent of teachers agreed on their survey that such improvements come about by empowering them to have a more prominent role with the new evaluation system while 100% agreed that the new system also encourages them to have more opportunities for reflection (Survey). In fact, reflection and collaboration were the most commonly cited positive effects of the evaluation system from the online survey.

Furthermore, a few changes in teacher behavior after the incorporation of the new evaluation system can also be cited as evidence of impact. For instance, the principal pointed out that he notices more teachers are volunteering for extracurricular activities than in the past (Participant A). A veteran teacher in the focus study also noticed this change; he thinks it is because the last Danielson component addresses this practice (Participant I). Additionally, the lead teacher pointed out that teachers are working together more often with the use of evaluation tools such as peer coaching (Participant C). Teachers from both focus groups agreed (Participant E; Participant F; Participant H; Participant I; Participant G) and further pointed out that these collaborative opportunities served as a formative assessment on how to specifically improve their practice especially since co-teachers have “real knowledge of the kids in my class and their strengths and weaknesses and their attitudes toward school” (Participant H). Moreover, such feedback
was used by teachers to change their practice (Participant E; Participant H). For instance, artifacts in portfolios showed teachers incorporating practices they learned in professional development training into their classroom (Archive D; Archive E). This trend was further made evident in the classroom observations where teachers referenced specific strategies they incorporated into their lesson plan and carried out in their classroom practice (Archive E; Archive F; Archive G).

With such opportunities for teachers to collaborate, reflect, and change their teaching practice, teachers are more likely to improve their teaching and therefore positively influence student achievement. For instance, the eleventh grade English teachers got together last year and completed a lesson study focused on grammar. They took part in peer review sessions and even critiqued classroom practice through video recording. The teachers then noticed that this year their English scores on the ACT were significantly higher, particularly the sub-score for grammar (Participant A, principal; Participant E, teacher). And as a high school, the campus cited overall improvements in ACT scores from the previous six years (Archive H). English and Reading had the highest percentage of proficient students ever and the highest average score. Science average scores were the highest in the past six years. And Math tied for second highest number of proficient students and increased almost half a point on its average from last year. It is unclear if the rise in test scores are directly correlated to the new teacher evaluation system. In fact, the central office representative and principal pointed to other reasons for the improvement in scores such as changes in curriculum (Participant B; Participant A). Nonetheless, they both agreed that the increased focus on teachers and their practice has played a role in improving student achievement. As one teacher said,
the new evaluation system “raises the bar essentially. It forces you to now, as a person, to raise your personal bar” (Participant E).

Case School #2

Background of the School

The middle school campus is located in a charter township of semi-rural neighborhoods, connected to a large metropolitan city with a history over 200 years. The township has a population close to 9,000 in the 2000 census with an average median income for a household of $44,277, which is slightly below the national average. The middle school’s mission is “to provide the academic growth and the skills needed to make positive choices in ensuring students become responsible, independent, and productive citizens in a global society” (School website).

The campus has an enrollment of 503 students from sixth to eighth grade. Within this student population, 39% are eligible for free or reduced lunch, which is close to the state average. Over ninety percent of student’s ethnicity is white. The school follows a two semester schedule with seven class periods of fifty-one minutes. Staff includes 26 teachers, 2 administrators, and 4 para-professionals. The principal was hired a year ago, after serving as the teacher union president. She shares operational and instructional responsibilities with her assistant principal, who was also hired a year ago. Table 13 shows detailed demographics of the school.
Table 13

Demographic Profile of Middle School (2011-2012)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grades served</td>
<td>6&lt;sup&gt;th&lt;/sup&gt; to 8&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Students served</td>
<td>503</td>
</tr>
<tr>
<td>Student population breakdown</td>
<td>White, not Hispanic: 92%</td>
</tr>
<tr>
<td></td>
<td>Black, not Hispanic: 5%</td>
</tr>
<tr>
<td></td>
<td>Hispanic: 3%</td>
</tr>
<tr>
<td></td>
<td>Asian/Pacific Islander: &lt; 1%</td>
</tr>
<tr>
<td></td>
<td>Native American/Alaskan Native: &lt; 1%</td>
</tr>
<tr>
<td>Students eligible for free or reduced-price lunch program</td>
<td>39%</td>
</tr>
<tr>
<td>Number of full-time administrators</td>
<td>2</td>
</tr>
<tr>
<td>Number of teachers</td>
<td>26</td>
</tr>
<tr>
<td>Per-pupil spending</td>
<td>$7510</td>
</tr>
</tbody>
</table>

Goal of the Teacher Evaluation System

The essential goal of the middle school’s evaluation system is focused on improving teacher practice through the use of the Danielson framework. This focus was emphasized by the central office representative, who saw the evaluation model as a way to “establish what is quality teaching” (Participant K); the principal also spoke of the system’s goal as primarily focused on “growth” (Participant J). Additionally, the lead teacher envisioned the system being “used as a tool to help teachers make improvements” (Participant L), while teachers in the focus group saw the system as providing opportunities and resources to be a better teacher (Participant N; Participant O). These reflections point to the primary purpose of the evaluation system, as written in the
evaluation manual: “to promote ongoing professional growth and performance which improves student learning” (Archive I). Such growth and performance improvement are designed to be provided for all levels of teacher experience through the use of the Danielson framework (Participant L, lead teacher; Participant J, principal). The ultimate hope is that such an evaluation system “provides the community with the assurance that their students are being taught by teachers of the highest quality” (Archive I).

Implementation of the Teacher Evaluation System

History. The current evaluation system was developed during 2008-2009 school year and then went into effect the following year. The previous system had been in effect for over 20 years. In 2007, the newly hired curriculum director was given the task of reviewing the evaluation model to measure its merit and impact on teachers. Flaws were evident from his review since the old system was strictly a checklist of items that the evaluator filled out once every three years. These checklist items consisted of less relevant questions such as “how the bulletin boards in the classroom looked” (Participant J, principal), “how you were dressed” (Participant L, lead teacher), and as a couple of teachers recalled with offense, it assessed if “the teacher is physically fit and able to withstand the stamina of the job” (Participant M; Participant R). Consequently, the director knew that a new system had to be developed. The main catalyst for beginning this process was dialogue from the state that possible changes in the evaluation system were being considered (Participant J, principal; Participant L, lead teacher; Participant M, teacher). As the principal stated, these reports were what “pushed us off the cliff” (Participant J) to begin the process of developing a new evaluation model.
The district formed an evaluation committee in 2008 with a team of 12 individuals, consisting of five administrators, the teacher’s union president and vice president, and five teachers from a variety of levels and subjects. After 14 months of meeting at least once a month for two to three hours each, the committee developed the new system with a pilot period for the five teachers in the evaluation committee that lasted a quarter of the school-year (Participant K, central office). After the pilot period and some modifications made to the system, the curriculum director presented it to staff members the following year, with full implementation beginning in the school year of 2009-2010. Then with the change in state law, the data portion of the evaluation system was added as the fifth domain during the second year of full implementation (Participant K, central office; Participant J, principal; Participant L, lead teacher).

**Description of teacher evaluation system.** The evaluation process for all teachers is similar, except for an individual development plan that is provided for new and first year teachers and teachers who received a less than satisfactory performance evaluation. The evaluation cycle begins with a teacher training of the evaluation system by the district, which is required for new and first year teachers. For the years thereafter, this training may be waived with teacher discretion.

The main portion of the evaluation system is the classroom observation. A minimum of two observations, each of 20 minutes or greater, is required during a school year for all teachers. One of these observations must be pre-arranged and additional observations may be announced or unannounced. During these observations, the evaluator will complete the Lesson Observation Report structured around the Danielson Framework (Appendix K). It should be noted that for first year teachers and teachers
who are in the development plan, Danielson’s domains two and three are only examined by the evaluator. After each classroom observation, a post-observation conference must be held within five school days. The post-conference will be structured around the review of two documents: the Lesson Observation Report and the Teacher Self-Checklist (Appendix L), which will be completed by the teacher beforehand for self-reflection. Additionally, classroom data is also discussed in this meeting as the fifth domain in the Lesson Observation Report. At the end of each post-conference, the teacher submits both forms after the evaluator and teacher discuss and review them together. If the teacher believes that the observation report is incomplete or unjust, they have the right to put these objections in writing, which is attached to the observation report. The first post-observation meeting and paperwork is due prior to December 15th.

Subsequent post-conference meetings that take place after classroom observations are meant to “access the amount of growth that has occurred” (Archive I). Similar to the first meeting, the evaluator and teacher will review the recently completed Lesson Observation Report and also the Teacher Self-Checklist from the first post-conference meeting. Any additional information to help in rating each domain of the Lesson Observation Report may be presented by the evaluator or teacher during these meetings. At minimum, it is required that a second observation and post-conference paperwork is submitted prior to March 15th.

The final step of the evaluation process is the summative evaluation meeting, where the evaluator will rate and review with the teacher all five domains using the Summative Evaluation Form (Appendix M). For these five domains, the summative form specifically rates teachers in each indicator of the Danielson framework into four rating
standards: exemplary performance, satisfactory performance, needs improvement, or unsatisfactory performance. These ratings are based upon the Danielson rubric (Appendix N). At the end of each domain is also a comment section, which the evaluator completes to provide evidence for their ratings. The fifth domain focuses on student performance data, which examines if the teacher has constructed assessments that show measurable and significant growth and if that data was used to modify instruction and assessments. The data is based upon growth percentage from beginning of year to end of year from national, state, and local assessments (Archive I). Thirty percent of the data component comes from NWEA’s national Measures of Academic Progress (MAP) assessment. Five percent is based upon state tests such as MEAP or MME, and fifteen percent will come from local assessments created by the district. A breakdown of these percentages is shown in Table 14.

Table 14

Middle School Evaluation System: Data Component

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Assessment (NWEA)</td>
<td>30%</td>
</tr>
<tr>
<td>State Assessment (MEAP or MME)</td>
<td>5%</td>
</tr>
<tr>
<td>Local Assessment (District-Generated)</td>
<td>15%</td>
</tr>
</tbody>
</table>

The Summative Evaluation Form breaks down the ratings from these five domains into two equal sections of 50% each, which includes the Danielson framework and the data component. These ratings are then calculated together in order to obtain an overall teacher evaluation rating. This overall performance range falls into four categories: exemplary for 90-100%, satisfactory for 75-89%, needs improvement for 60-
74%, and unsatisfactory for 0-59%. This overall rating will determine the final employment recommendation, which will be submitted at the end of this form as either: probationary status, tenure, or non-renewal. A recommendation for tenure is made if the overall evaluation is rated as “exemplary” or “satisfactory”. Ratings of “need improvement” will lead to probationary status while “unsatisfactory” rating leads to non-renewal. Any objections can be submitted and attached to this document in writing by the teacher. In such cases, the evaluation form will be “reviewed during the school year by representatives of the Board and the association to determine its validity and changes may be made as a result of such review” (Archive I).

As mentioned earlier, new and first year teachers and teachers who have received a less than satisfactory performance evaluation are required to complete an Individual Development Plan (Appendix O). This plan is completed by the evaluator in consultation with the teacher. The form is structured around forming and monitoring specific goals that are “measurable, behavioral objectives” with specific actions and outcomes identified (Archive I). The first two goals are focused on indicators from domain two of classroom environment and domain three of instruction from the Danielson framework. Specifically, goal one is “to develop instructional delivery skills for providing student success” while goal two focuses on providing “sound classroom management that is conducive to learning” (Archive I). Unlike the standard evaluation system, it should be noted that teachers who fall under the Individual Development Plan are evaluated only on these two domains from the Danielson framework. To accomplish the first two goals for beginning or struggling teachers, the development plan requires that the evaluator and teacher build a specific action plan on what the teacher will do and also an outline of the
Administrative support provided for teachers. Then the third goal in the development plan is concretely tied to how beginning teachers will meet 15 days of six hour durations of professional development training over the first three years of teaching. There is also an optional fourth goal that can be developed by the teacher and evaluator upon mutual agreement. The Individual Development Plan is completed and reviewed during each post-observation meeting. Finally, during the summative meeting, the Individual Development Plan is used to make an employment recommendation decision, submitted through the Summative Evaluation Form.

Figure 4 provides a flowchart of which domains of the Danielson framework are evaluated based upon teacher classification.

Figure 4. Flowchart of two tracks of middle school evaluation system.
Survey Results: Case School #2

For the second school’s online survey, 16 teachers submitted responses out of 26 teachers in the middle school campus, for a response rate of 61.5%. As mentioned earlier, majority of experts in the field recommends a 60% response rate for generalizability of results (Glatthorn, 1998), which was achieved among teachers from this first case study.

Specific response numbers and percentages are shown in Tables 15-18. First, Table 15 provides a breakdown of the initial two questions in the survey where teachers were asked to indicate amount of classroom teaching experience at current campus and also total number of teaching experience. This table shows number of teachers considered new teachers (those with five years of experience or less) compared to veteran teachers. Results show a majority of teachers have five years or less of experience at the middle school campus.

Table 15

*Distribution of New and Veteran Teachers: Case School #2*

<table>
<thead>
<tr>
<th></th>
<th>Number of Teachers with Years of Teaching at Current Campus</th>
<th>Number of Teachers with Total Years of Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Teacher</td>
<td>11 (69%)</td>
<td>5 (31%)</td>
</tr>
<tr>
<td>Veteran Teacher</td>
<td>5 (31%)</td>
<td>11 (69%)</td>
</tr>
</tbody>
</table>

The next portion of the survey included 19 Likert graded statements. Tables 16-19 display percentages of the 16 teachers who provided responses in the Likert scaled...
section of the survey, with statements that are condensed from the original survey and categorized according to the four research questions.

Table 16 shows respondent’s Likert scale percentages for statements under fairness, usefulness, and accuracy categories that relate to the first research question about the makeup of the evaluation system that is based upon the Danielson framework. Majority of teachers agreed that the Danielson framework provides an accurate representation of teacher practice. Additionally, most teachers agreed that the model is useful in providing opportunities for input and self-reflection. On the other hand, results were mixed as to the overall fairness and efficiency of the model demonstrating teacher practice, and majority of teachers disagreed that the model provides adequate number of data sources to offer a richer and comprehensive view of teacher practice.

Table 16

*Statements Categorized Under School Addressing Danielson’s Domains (Research Question #1)*

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair and Efficient</td>
<td>6%</td>
<td>31%</td>
<td>31%</td>
<td>19%</td>
<td>13%</td>
</tr>
<tr>
<td>Prominent Role</td>
<td>13%</td>
<td>56%</td>
<td>0%</td>
<td>19%</td>
<td>13%</td>
</tr>
<tr>
<td>Two-Way Communication</td>
<td>38%</td>
<td>44%</td>
<td>13%</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>Promotes Good Teaching</td>
<td>19%</td>
<td>63%</td>
<td>19%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Encourages Self Reflection</td>
<td>13%</td>
<td>81%</td>
<td>6%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Change Instructional Practices</td>
<td>0%</td>
<td>81%</td>
<td>19%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

174
Table 16–Continued

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify Strengths and Weaknesses</td>
<td>25%</td>
<td>56%</td>
<td>13%</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>Danielson Provide Accurate Representation</td>
<td>13%</td>
<td>63%</td>
<td>25%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Provide Evidence</td>
<td>0%</td>
<td>50%</td>
<td>38%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>Richer and Comprehensive</td>
<td>0%</td>
<td>50%</td>
<td>44%</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>Adequate Data Sources and Inputs</td>
<td>0%</td>
<td>31%</td>
<td>63%</td>
<td>6%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 17 shows respondent’s Likert scale percentages for two statements under feasibility category that connect with the second research question, “What resources are utilized to implement a multidimensional teacher evaluation system?” Over half of respondents provided agreement on these two survey questions. Specifically, 62% of teachers strongly agreed or agreed that evaluation system’s policies were clearly written and provided; 56% of teachers also agreed that resources for the evaluation model were sufficiently provided by the district.

Table 17

*Statements Categorized Under Use of Resources (Research Question #2)*

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearly Written &amp; Provided Policies</td>
<td>6%</td>
<td>56%</td>
<td>25%</td>
<td>6%</td>
<td>6%</td>
</tr>
</tbody>
</table>
Table 17—Continued

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sufficient Resources</td>
<td>0%</td>
<td>56%</td>
<td>38%</td>
<td>6%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 18 displays respondent’s Likert scale percentages for the first three statements under the feasibility category that connect with the third research question, “What challenges do schools face in implementing a multidimensional teacher evaluation system and how are they addressed?” Though a majority of teachers (75%) agreed that reasonable time is provided to review evaluations, most teachers with 63% felt that professional development and training on the evaluation system were not sufficient. On the other hand, a slight majority of teachers with 56% agreed that their evaluator had necessary training, knowledge, and skills to conduct evaluations.

Table 18

*Statements Categorized Under Challenges Encountered (Research Question #3)*

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Reasonable</td>
<td>0%</td>
<td>75%</td>
<td>25%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Teacher Training Sufficient</td>
<td>0%</td>
<td>31%</td>
<td>44%</td>
<td>19%</td>
<td>6%</td>
</tr>
<tr>
<td>Evaluators Trained</td>
<td>0%</td>
<td>56%</td>
<td>31%</td>
<td>13%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Lastly, Table 19 displays respondent’s Likert scale percentages for the final three statements of the survey that connect with final research question, “What evidence(s)
exists that the teacher evaluation system has or is resulting in positive educational outcomes?” While a majority of teachers (62%) disagreed with the statement that the former evaluation system helped contribute to their professional growth, results were mixed with about half agreeing or disagreeing as to the effect of the new evaluation model leading to student growth and appropriate professional development training.

Table 19

*Statements Categorized Under Evidence of Growth (Research Question #4)*

<table>
<thead>
<tr>
<th>Likert Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Former Model Provide Growth</td>
<td>0%</td>
<td>25%</td>
<td>6%</td>
<td>56%</td>
<td>13%</td>
</tr>
<tr>
<td>New Model Lead to Student Growth</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
<td>0%</td>
</tr>
<tr>
<td>New Model Lead to Appropriate Training</td>
<td>0%</td>
<td>38%</td>
<td>25%</td>
<td>25%</td>
<td>13%</td>
</tr>
</tbody>
</table>

The final portion of the survey included an optional free response section for respondents to provide their personal views regarding major advantages and disadvantages of the evaluation process. From the 16 teachers who submitted an online survey, seven teachers (44%) submitted their thoughts on the major advantages of the evaluation system. Additionally 10 teachers (62%) also provided their views on the major disadvantages of the evaluation system. In order to evaluate the free response section, each submission was categorized from the transcripts for word frequency and thematic phrasing.

After completing the above analysis, two major themes were present in the first section that relates to the major advantages of the evaluation process. The majority of the
comments from six teachers related to the evaluation model providing multiple domains that reflect various aspects of teacher practice. Comments ranged from “Very specific behaviors spelled out for you in the tables of domains” to “There are very specific domains and expectations” and “teachers are evaluated on multiple relevant criteria.” Two teachers’ comments were based upon the theme of reflection with comments such as “leads to more self-reflection” and “opportunity for reflection”.

Two major themes also came from the analysis of the second section that relates to the major disadvantages of the evaluation process. Four teachers commented that the evaluation system does not provide a complete picture of their practice. Comments from this theme ranged from “administration may only see parts of what you do in your classroom, not the whole picture” to “Only 2 observations to get a picture of all of the domains doesn't seem like enough” and lastly, it is “not a true picture of teacher performance.” The second major theme with three responses fell under issues regarding time with comments such as “It seems this process takes substantially longer to adequately meet the requirements” and “It takes an incredible amount of time to evaluate teachers.” As the later sections of this chapter will cover, these free responses are valuable in providing further support to data collected from documents, interviews, and observations.

Research Questions Examined

Question #1: Danielson framework. The first research question examines how the school is addressing each of Danielson’s four domains of teacher practice. In order to answer this question, it is first necessary to establish why the campus chose this framework.
As mentioned earlier, the purpose of the Danielson framework is to provide common definitions of expertise and procedures for effective teaching; “Because teaching is complex, it is helpful to have a road map through the territory, structured around a shared understanding of teaching” (Danielson, 1996, p. 2). When the middle school’s evaluation committee was formed, their first discussions were based around a similar need to find a common and shared understanding of teaching (Participant K, central office). The Danielson framework was chosen because it offered a systematic approach to evaluating teachers that is both specific and flexible to teacher practice (Participant K, central office; Participant L, lead teacher; Participant Q, teacher; Participant N, teacher; Participant O, teacher).

The first reason for selecting the Danielson framework is due to its specificity on teacher practice. The Danielson framework is organized according to four domains or realms of activity that refer to distinct aspects of teaching. Danielson (2007) identified the four domains as planning and preparation, the classroom environment, instruction, and professional responsibilities. Each of these teaching domains contains detailed components that identify a range of appropriate teacher behaviors, with rubrics to assess unsatisfactory, basic, proficient, and distinguished performance. Such specificity appealed to teachers in the focus group as well since it provides “a better understanding of what is expected of me” (Participant N) with a detailed rubric that outlines “how we can improve and get to the exemplary category” (Participant O). Furthermore, 76% of teachers agreed on the online survey that the Danielson framework provides an accurate representation of teacher practice. This was especially a refreshing change for those
teachers who took part in the previous evaluation system that focused on areas disconnected from teacher practice (Participant P; Participant R; Participant M).

The other reason for choosing the Danielson framework was its flexibility. The campus’ evaluation system makes use of the four Danielson domains depending on the specific stage of a teacher’s practice. New and first-year teachers and teachers who fall under the individualized development plan focus only on domains two and three during those evaluation periods (Participant J, principal). This differentiation ensures that these teachers are first developed in the primary skills of classroom management and instruction. It also provides both the evaluator and teacher a focus on more specific areas of growth. In other words, this differentiation allows them the opportunity to “pinpoint exactly where the teacher might need help” (Participant K, central office) so that the process is made easier to “zone in” (Participant N, teacher) and “pinpoint exactly what I’m doing” (Participant O, teacher) with regards to the classroom environment and instruction. And in doing so, the other two domains that fall outside the classroom practice, specifically planning and preparation and professional responsibility, will only be evaluated once the classroom environment and instruction are first established for new or struggling teachers (Participant K, central office; Participant J, principal). Once these teachers are rated at least satisfactory in these two domains, then they will be rated according to all four domains the following year.

In examining the evaluation system as a whole, the four Danielson domains are primarily addressed during the principal classroom observation and the teacher self-reflection process. The Lesson Observation Report (Appendix K) is built around these four domains with Danielson indicators labeled under each domain. Evaluators ensure
that all their comments and feedback are based on these four domains by noting them specifically under the appropriate section. The first domain of planning and preparing is evaluated by the principal after looking at the teacher’s lesson plans, which are submitted to the principal every week, and also after considering the comments made by the teacher in their teacher self-checklist (Participant J, principal). Then during the actual classroom observation, principals note their observations with the lens of the next two domains of classroom environment and instruction. Before determining if the indicators under those domains were observed, the principal also considers the Teacher Self-Checklist with teacher comments listed under the respective two domains (Archive J). In fact, teachers mentioned that the self-reflection process allows evaluator and teacher to converse about domains individually; they have opportunities for input provided to their evaluator by noting self-reflection scores for each indicator (Participant N; Participant R; Participant P). Furthermore, survey results also showed that 94% of teachers agreed that the new evaluation system promotes self-reflection while 82% agreed that it also promotes two-way communication between evaluator and teacher.

The last Danielson domain of professional responsibilities is primarily addressed by evaluating evidence provided by teachers or observed by administrator in the following five indicators: reflecting on teaching, maintaining accurate records, communication with families, growing/developing professionally, and showing professionalism. It is the teacher’s responsibility to bring evidence that connects with these five indicators, which is cited specifically in their Teacher Self-Checklist (Appendix L) under domain four. For instance, teachers can provide intervention logs as evidence of communication with families, which requires documenting at least two
different contact methods to reach out to parents (Participant J, principal). At the same
time, the principal also verifies evidence of professional responsibilities by accessing the
teacher’s gradebook and records of attendance taking to evaluate the indicator of
maintaining accurate records. Teachers also attach evidence of attending professional
development training in their Teacher Self-Checklist in order to address indicator of
developing professionally. And in observing post-conferences, one could see that the
principal addressed the indicators of reflecting on teaching and developing professionally
by citing evidence of recent professional development practices that were seen in action
in classroom observations. For example, one teacher’s questioning technique and
another’s reading apprenticeship training were cited in their Lesson Observation Report
under domain four (Archive K; Archive J). It should lastly be noted that when teachers
also cite these evidence in their Teacher Self-Checklist, they are also provided another
opportunity to reflect upon their practice as well.

**Question #2: Resources.** The second research question focuses on examining
how resources are utilized to implement the school’s teacher evaluation system. At the
beginning of the process, much of the resources were based on time required to form and
develop the evaluation model. The evaluation committee met at least once a month for
fourteen months for two to three hours each. Since these members were volunteers, there
were not any major monetary costs involved for the district. Time was also the major
resource for implementing the new evaluation system. For instance, administrators were
the main personnel who spent their time in conducting the evaluations. This
responsibility was in addition to their job requirement of maintaining the operation and
instructional programs of the school. Another prominent staff member who helped
implement the system was the lead teacher who served as resident expert whom teachers can ask questions and express concerns. Additionally, five teachers were involved in the pilot program for a quarter of the school year. But it should be noted that the lead teacher and pilot teachers served in their capacities as volunteers. Finally, all staff members who took part in the full evaluation process used their time during and outside school hours to complete their self-checklist, meet with their evaluator, and collect evidence related to the five domains.

On the other hand, there were some direct costs that the district had to bear to develop and implement the evaluation system. One major budgetary cost was trainings provided to the central office representatives, administrators, and teachers. Initially, the central office representative was trained in the Danielson framework. He then shared this training with his administrators and teachers. These trainings took up a major portion of the two and a half days of staff training given at the beginning of the school year (Participant K, central office). Then throughout the school year, ongoing trainings were provided as needed to administrators and teachers on various elements of the evaluation system. For instance, the middle school principal noted that she used about two to three hours of staff training time interspersed throughout the school year to focus solely on the evaluation system (Participant J).

An additional cost for the district is computer programs that were purchased to help implement the evaluation system. One of these programs is Data Director that serves as the data warehouse for the campus with all evaluation data, teacher gradebook, and classroom assessments entered into this system (Participant K, central office; Participant J, principal). The program then allows central office representatives,
administrators and teachers to reconfigure and analyze the data depending on the user level, with tables and charts. The second program purchased through NWEA is the online assessment called Measurement of Academic Progress (MAP) that is used for pre-and post-assessment data. This assessment is given to students twice a year. The system provides a pre- and post-score, which is included as the national assessment in domain five of the evaluation system. Both of these systems require computers, which the district already provides to every teacher, while students receive computer access through the school’s computer lab. These computers are also required by administrators and teachers to access and complete their evaluation forms as well.

**Question #3: Challenges.** The third research question examines the challenges schools face in implementing their teacher evaluation system and how they are addressed. From the central office representatives to administrators and teachers, the middle school faced a number of challenges that were addressed in various degrees. These challenges include staff buy-in, professional development and training, and time constraints.

**Professional development and training.** The most common challenge expressed by administrators and teachers was concerns about the professional development and training provided for all stakeholders. In fact, the administrator, lead teacher and staff members in the focus group all expressed concerns about insufficient training provided to teachers about the new evaluation system (Participant J; Participant L; Participant M; Participant S). Survey results attest to a similar conclusion since 63% of teachers disagreed with the statement that the professional development and training on the evaluation system for teachers were sufficient. The central office representative admitted that though the district offered three sessions at the beginning of the year to review the
new evaluation system, only two teachers participated in these optional trainings (Participant K). A major reason for this lack of participants was due to lack of buy-in from staff members, which will be covered in the next section. Another major reason was due to miscommunication between the district and teacher association on who would provide these trainings; one side assumed that the other would train the teachers (Participant J, principal; Participant L, lead teacher; Participant T, teacher). Additionally, the middle school campus experienced a number of turnovers in administration and teachers, which complicated this issue even further. For instance, both the principal and assistant principal joined the campus only this year. Plus, the campus went through three principals in the past three years (Participant J, principal; Participant P, teacher). It is no wonder then that staff members posed questions about the evaluation model during interviews. These questions ranged from not knowing as much information about the four Danielson domains (Participant S; Participant P), on how to use the Danielson rubric (Participant M; Participant S; Participant Q; Participant R), and fears about how they would be evaluated (Participant M; Participant Q).

The middle school campus has attempted to address these concerns with ongoing training on the evaluation system throughout the year (Participant J, principal). Since the campus is attempting to fill as much of the gap as possible with the limited time left in the school year, these trainings have particularly focused on areas of highest concern. For instance, the principal noted in her classroom observations that the Danielson indicator of questioning techniques was lacking amongst her staff members. Therefore, her recent training focused on domain 3, indicator 3b, which examines using questioning and discussion techniques. Staff members examined this domain and indicator together
and then watched an example of how to more effectively use questioning skills in the
classroom. The teachers ended the training by rating a videotaped classroom lesson using
the Danielson rubric. The principal understood that these types of comprehensive
trainings are needed on an ongoing basis for her staff members. Teachers also responded
positively to this training because they saw value in seeing specific classroom examples
of how to put into practice the Danielson domains and what their evaluators are looking
for when observing their classrooms (Participant R). The principal hoped to continue
these types of trainings at the campus and district level; she admitted, “I think we should
have spent some time doing it [training] a little bit more in-depth with the teachers”
(Participant J).

**Staff buy-in.** Another major challenge for the district was concerns about staff
buy-in of the new evaluation system. It should first be mentioned that some teachers
expressed positive feedback in regards to the collaborative nature in which the system
was developed (Participant M; Participant O; Participant L). Out of the twelve member
evaluation committee, teachers were represented by seven members that included the
teacher’s union president and vice president and five teachers from a variety of levels and
subjects. Initially, time was spent during those fourteen months of building the
evaluation system to discuss and build trust amongst the committee members (Participant
K, central office; Participant J, principal; Participant L, lead teacher). In fact, the central
office representative cited “communication, trust building, and having a variety of people
around the table” as reasons for successful development of the evaluations model
(Participant K). These members also bought into the change because they understood
that changes were coming from the state; therefore, by forming a system locally, the
district would have a greater voice in how their teachers are evaluated (Participant L, lead teacher; Participant J, principal; Participant K, central office). On the other hand, when the evaluation system was presented to the staff, this collaborative approach was missing. The central office representative presented the new system district-wide in a meeting led by only him. In doing so, staff members could assume that this system was developed centrally (Participant J, principal). The central office representative admitted to this mistake when he said, “When presented to the group as a whole, something like this, nobody stands alone on the stage . . . if we collaborate to put it together, it has to be a collaboration in presentation” (Participant K). Subsequently, staff buy-in of the new evaluation system was negatively impacted.

At the same time, a lack of consistent leadership at the central office and campus level could be another factor that influenced the system’s lack of acceptance amongst teachers. For instance, during the fourteen months in which the evaluation system was being developed, the district went through three superintendents (Participant K, central office), which most likely contributed to lack of consistent communication about the evaluation system. And at the local level, the middle school also went through three principals in the past three years (Participant J, principal; Participant P, teacher). Hence, the local expertise in the new system was lacking as well (Participant K, central office). These ongoing changes in leadership would most likely have contributed to teachers not accepting the new system. In their mind, they could have perceived this new system also changing with changes in leadership (Participant P, teacher); in other words, some teachers may have held the “disbelief that there was going to be that change [in the evaluation model]” (Participant K, central office).
In examining how the district addressed this concern of staff buy-in, the consensus was that it was not effectively addressed (Participant J, principal; Participant K, central office). Training was offered to teachers after the new system was presented, but most teachers did not attend. As the central office representative learned, when implementing such a change, it is important to “take your time, do it right once, get buy-in once” (Participant K); if these initial steps are not taken, then the subsequent challenges are harder to address later on. Once the system was implemented and teachers were evaluated by it, staff members began to ask for more training (Participant K, central office; Participant M, teacher). As the principal mentioned, “I don’t know we did effectively address it. It has become over time the way we do things.” (Participant J)

The challenge of staff buy-in of the evaluation model was especially evident with the data component of the model. This lack of acceptance of the data component was not just seen by teachers but also from the lead teacher who was a part of the original evaluation committee. In fact, she referenced the data component in the following manner: “I don’t see the student data as a necessary part of it [evaluation system]; more of an unfair part of it” (Participant L). One major reason lies in the fact that the data component was added later to the district’s evaluation system as the fifth domain because of state mandate (Participant L, lead teacher; Participant J, principal; Participant K, central office). The other major reason for lack of staff acceptance is their view that connecting student data to teacher evaluation is unreasonable because other factors affecting student achievement is not considered (Participant L; Participant N; Participant T; Participant P). These other factors range from “parent component” (Participant L) to “learning difficulties” (Participant T) and “even prior knowledge” (Participant O). As
one teacher summarized by saying, “there’s a lot of factors” associated with student achievement such that “there is a lot more than me” as the teacher that should be considered (Participant P).

In order to address these concerns, the district developed the evaluation model with certain criterion. First, no one data is used exclusively in the data component. As detailed earlier in Table 10, domain five of the evaluation model is based upon growth percentage from beginning of year to end of year from national, state, and local assessments (Archive I). Thirty percent of the data component is taken from NWEA’s national Measures of Academic Progress (MAP) assessment. Fifteen percent comes from local assessments created by the district and the last five percent is based upon state tests such as MEAP or MME. State tests only account for five percent because these tests currently do not show actual growth data since pre-tests do not exist and the state data from one year to the next are based upon different outcomes (Participant K, central office). To fill this gap, the district developed local pre- and post-unit assessments. The central office representative admitted that creating quality assessments is actually an ongoing work in progress with district-wide professional development focused on how to write quality assessments and how to interpret the data from these assessments to guide classroom instruction. In addition to including a mix of national, state, and local assessments, the data component also is set up to look at three year trends. This avoids making teacher evaluation decisions from only one year of student data results. And these results are based on cohort of students and not individual students. Lastly, it is important to note that the new evaluation system with the data component is also used when evaluating administrators and central office representatives. As the central office
representative clarified, teacher data is also attached to his data “because I’m evaluated by student data also” (Participant K). As such, he must ensure that the system is reliable since he is also evaluated by it. These various conditions in the data component of the evaluation system attempts to address previously mentioned reliability concerns from teachers.

**Time constraints.** The final challenge expressed by both the administrator and teacher is the extra time commitment necessary in implementing the new evaluation system. This challenge began at the initial stage with the evaluation committee, which met at least one time each month for two to three hours, lasting for a total of fourteen months before an evaluation model was developed. After the pilot period and some modifications made to the system, the curriculum director presented it to staff members the following year. It is important to note that such a time commitment is required when developing a new evaluation model.

Once the model was implemented in the middle school campus, teachers expressed concerns with finding time to balance their teacher duties with the additional responsibilities from the evaluation system. For instance, teachers are now required to document all their professional responsibilities under domain three, which can be “time heavy” (Participant M). Another teacher expressed concerns about collecting data for the data review component of domain four which will now require her to start “spending a lot of time proving that I am actually a good teacher” (Participant T). At the same time, other teachers noted that collecting information for the evaluation meetings took less time the second time around since they were more aware of what’s required and therefore could plan accordingly (Participant N; Participant M; Participant O). Hence, the
challenge of time is not necessarily a “bad thing; it’s just a different challenge that I’m not used to from before” (Participant N).

The administrator also expressed concerns about finding time to evaluate her 30 teachers every year with “two observations a piece, pre-observation, and post-observation; there’s a lot of meetings” in addition to her other responsibilities (Participant J). To address this challenge, she mentioned that she now “works smarter” (Participant J) by having teachers use Google Docs to keep their records. She also has learned to use her meeting time more efficiently. And lastly, she plans to provide her new assistant principal more responsibility in evaluating teachers next year after he is trained in the evaluation model.

It should also be noted that the evaluation system is different for new and struggling teachers, with only domains two and three evaluated by administrators. Furthermore, the central office representative envisioned the evaluation system becoming further differentiated in the future so that experienced teachers, who have proven to be highly effective in the classroom, would not be observed by administrators but would rather assistant new and struggling teachers (Participant K). These types of differentiated evaluations would further address the challenge of time for both teachers and administrators.

**Question #4: Evidence of impact.** The final research question examines what evidence exists that the teacher evaluation system has or is resulting in positive educational outcomes. Most respondents agreed that there are not concrete data that show positive impact for the middle school after the implementation the evaluation model. However, this is due primarily to the infancy of the model since it has only been
in existence for two years (Participant J, principal; Participant K, central office).

Moreover, the middle school campus went through administrator changes the past three years during the development and implementation stages of the evaluation model, which further impacted the level of continuity in using the new system (Participant O, teacher; Participant K, central office); the principal in fact commented that “the first year [the evaluation system] was being used, it was so new that I’m not sure that administrators fully implemented it” (Participant J).

Administrators and teachers nonetheless pointed out that the evaluation system influenced the overall climate of the middle school campus. For instance, the most common feedback about the system’s impact is reflected by the comment from one teacher that “it’s a different climate here now” (Participant M); this climate change specifically focused on greater accountability for teachers. Teachers now have a tool with the Danielson framework to examine specific areas of growth. This tool is used to improve practice for all teachers since “there’s always going to be something that I feel that I can always improve on” (Participant L, lead teacher). Therefore, all teachers from beginners to veterans are being held more accountable. With the Danielson rubric that lists the four domains and specific indicators under each domain, teachers responded that they have a better idea of what’s expected of them with specific areas to improvement in teacher practice (Participant N; Participant O; Participant P; Participant R; Participant L). These expectations combined with the Teacher Self-Checklist also led to more teacher self-reflection on what they are teaching and how it impacts students (Participant M; Participant O; Participant N; Participant P). The principal also echoed similar comments,
where she observes that teachers are beginning to look “more critically at what they’re doing” (Participant J).

The climate of accountability positively and negatively affected teachers in regards to the data component of the evaluation system. On one end, the tracking and use of data has shifted conversations to a focus on numbers and results. Evaluations have become “more objective” (Participant J, principal) and “less subjective” (Participant M, teacher) so that “it’s not just the opinion of the principal” (Participant K, central office). There is more emphasize on meeting growth goals for students since domain five is built around the components of demonstrating appropriate student achievement and academic growth while also using data to modify instruction and assessments (Archive I). Teachers have responded that their focus is more tailored towards meeting the established academic growth goals in their classroom (Participant O; Participant M; Participant T). For instance, the principal spoke about data meetings tied to conversations around pre- and post-test data with a detailed analysis of assessments in order to develop an intervention plan (Participant J). Documents from post-conferences also attest to these types of analysis with detailed question by question examinations of assessments (Archive J; Archive K; Archive L). On the other hand, teachers also expressed negative implications from the data component as well. As mentioned earlier, these teachers felt that connecting student achievement to teacher evaluation is unreasonable due to the numerous outside factors that impact their students (Participant L; Participant N; Participant O; Participant T; Participant P). Therefore, they expressed reservations about the reality that half of their evaluation scores will be tied to student data (Participant N; Participant M; Participant T).
Additional impacts from the climate of accountability are specific programs that began with the institution of the new evaluation system. For instance, the middle school campus instituted a homework tracking system this year to avoid zeros in assignments so that the failure rates are reduced (Participant P). Moreover, teachers are now required to keep track of failing students with a log and documentation of steps taken to mitigate the problem. These steps tie specifically into domain four of the evaluation system of professional responsibility.

It is important to note that the climate of accountability extended to administrators and central office personnel as well. With ongoing observation data across various domains of teaching, the district and campus leadership are examining this data to find common areas of development for teachers. They then “develop professional development to address those issues” (Participant K, central office). For instance, at the district-level, trainings were focused on developing classroom management skills since these skills were seen as a deficiency for many teachers. At the local level, the middle school principal saw the need to lead trainings focused on the art of questioning that tied into domain 3, indicator 3b, which examines using questioning and discussion techniques. These steps further attest to the impact of greater self-reflection from all stakeholders within the district so that “we’re constantly reviewing where we are at and how to get better” (Participant K, central office).
CHAPTER V
DISCUSSION AND CONCLUSION

Previous research studies have shown the vital role that teachers play in influencing student achievement (Borman & Kimball, 2005; Darling-Hammond, 1997b). With this understanding, federal, state, and local leaders have been examining ways to hold teachers accountable for their share in meeting student growth goals. Subsequently, one of the methods identified in this study to hold teachers accountable by measuring teacher performance is via an effective teacher evaluation system; after all, Stronge and Tucker (2003) wrote, “without high quality evaluation systems, we cannot know if we have high quality teachers” (p. 3). On the other hand, the first chapters of this study showed deficiencies in research literature related to the process around implementation of an effective teacher evaluation system. Hence, the purpose of this study is to partially fill this void by examining strategies used by two Michigan schools to build research-supported multidimensional teacher evaluation systems. This final chapter will provide a synthesis of the results obtained from the case studies at the two schools. The chapter is divided into three sections: discussion of the results, recommendations, and ideas for future research. These sections will revolve around the four research questions with a primary focus on strategies and processes used by these two schools. After all, the practical value of this study is to identify and describe promising practices of a multidimensional teacher evaluation system based on the Danielson framework.

Discussion of Results

The examination of results from this study uncovered similarities and differences in the development and implementation of the teacher evaluation system between the two
schools, as well as connections to previous research on this topic. This section will look into these comparisons as it relates to the four research questions.

**Research Question #1: Danielson Framework**

The two schools chose the Danielson framework because they believed that the domains of this framework provide a common definition of teacher expertise that is needed in order to evaluate teacher practice (Participant A, principal; Participant C, lead teacher; Participant K, central office; Participant L, lead teacher). In fact, teacher interviews and survey results from both schools indicated that this framework offered a common standard for describing effective teacher practice with rubrics to assess performance (Participant R; Participant N; Participant O). Both schools saw the value of such a framework when it was compared to their previous evaluation model that focused on areas disconnected from teacher practice (Participant A, principal; Participant B, central office; Participant C, lead teacher; Participant P, teacher; Participant R, teacher; Participant M, teacher). These findings therefore reiterate previously mentioned studies that show the efficacy of using the Danielson framework within a teacher evaluation system (Heneman, Milanowski, Kimball, & Odden, 2006; Holtzapple, 2003; Kimball, White, Milanowski, & Borman, 2004; MET, 2012; Sartain, Stoelinga, & Krone, 2010; Taylor & Taylor, 2003; Toch & Rothman, 2008).

The two schools examined the four domains of the Danielson framework through their various evaluation instruments used to collect data. For instance, both campuses primarily made use of the administrator’s classroom observations to collect evidence of their teachers addressing the Danielson domains. Then, depending upon the year of teaching, teacher choice, and previous evaluation status of the staff member, the high
school campus used one of the following instruments to collect additional data: teacher portfolio, peer coaching, lesson study, video lesson, or action research. For the middle school, besides the lesson observation report that is filled out by the administrator, their evaluation system used teacher self-checklists and individual development plans to address the Danielson domains. With the use of these various evaluation instruments, the model at both schools becomes multidimensional.

The second portion of this research question asks how the multidimensional evaluation systems help bring about transformation in the schools. Teachers from both schools acknowledged that these various instruments used to collect data of their teaching practice allowed them to play a more prominent role in the evaluation process, while also encouraging self-reflection and collaboration (Survey; Participant C; Participant E; Participant F; Participant H; Participant I; Participant G; Participant M; Participant O; Participant P). These end-results are important because previous researchers have showed that these same practices of collaboration and self-reflection are key propositions that assist in identifying and recognizing effective teachers (Abbott & McKnight, 2010; Airasian & Gullickson, 1994; Blasé, 1987; Darling-Hammond, 1998; National, 1989;).

On the other hand, the two schools’ guideline regarding the number of teacher observations using the Danielson framework is limited. Specifically, the administrators at both schools are only required to observe their teachers two times within a school year. For the middle school, this requirement is further shortened with a minimum of only 20 minutes per observation versus a full class period for the high school. Subsequently, teachers surveyed from the middle school in particular commented that the evaluation system does not provide a complete picture of their practice; as one teacher wrote, “only
2 observations to get a picture of all the domains doesn’t seem like enough”. In fact, previous research has shown that when more frequent observations occur, the reliability of these observations improve (Denner, Miller, Newsome, & Birdsong, 2002). Thus, I believe that leaders at both districts should look at ways to increase the minimum number of evaluation feedback with their teachers. When investigating these options, the districts can look into ways to incorporate multiple evaluators or evaluation instruments to address the Danielson domains so that principals are provided with multiple perspectives while they can also share evaluation responsibilities with others. For instance, the principals at both schools can utilize their assistant principal to help observe teachers, after the assistant principal has obtained sufficient training. This additional step will provide further feedback for teachers from different evaluators while also serving as an opportunity to train and develop their assistant principal. Besides, previous researchers have already shown the advantages of a multidimensional evaluation system that make use of multiple and variable data sources (Darling-Hammond, 2006a; Odden, 2004; Peterson, 2006; Stronge, 2006).

**Research Question #2: Resources**

Both schools used similar resources in developing and implementing their evaluation system. Time was the first major resource necessary to develop the system. Initially, considerable time was needed to form and then meet regularly with the evaluation committee. Then time was spent piloting the program, training staff members, and finally implementing the new evaluation model. For instance, classroom observations by administrators played a significant part in the new evaluation system at both sites. The high school principal spent approximately 30% of his time observing and
evaluating teachers, which is equivalent to 2 hours a day. At the middle school, the principal explicitly mentioned the increased time commitment necessary to observe her 30 staff members once the new evaluation model was implemented. On the other hand, it should be noted that the increased use of time also allowed these schools to focus more resources on instructional practices. According to prior research, the allocated resource of time for classroom and teacher visits allow the principals, as instructional leaders, to monitor their instructional program, provide teacher feedback, and support teacher development (Blase, 1987; Hallinger & Heck, 1998; Heck, Larsen, Marcoulides, 1990; Reitzug, West, & Angel, 2008; Siens & Ebmeier, 1996). These practices, after all, are important steps in meeting the goal of increasing student achievement.

As far as budgetary resources, both schools did not have major monetary costs throughout the evaluation process since the committee, administrators, and teachers were not given additional pay to develop and implement the system. However, some direct costs were initially needed to hire consultants to train staff. Additionally, both schools had to purchase a data program and the standardized NWEA online assessment in order to measure student achievement growth.

On the other hand, there was a common criticism from both schools about the inadequacy of resources related to the policy and guideline around the new evaluation system. For instance, teacher interviews and survey results showed that staff members felt that the policies and guidelines of the new evaluation system could have been more clear. Teachers at the high school in particular commented that they were confused about the various evaluation instrument options given to them (Participant C; Participant G; Participant H). It will certainly be important for both schools to address this concern
since research shows that adequate and clear policies is one essential attribute necessary for a fair and sound evaluation system (Brandt et al., 2007; Joint, 2009; Koppich & Showalter, 2008; O’Day, 2002). Details on how to possibly address this concern will be covered in the next section of this chapter.

**Research Question #3: Challenges**

The common challenges that both campuses encountered when developing and implementing their evaluation system is staff buy-in and time constraints. The middle school had challenges with staff buy-in of the new evaluation system partially due to ongoing changes in leadership with their principal and superintendent. Earlier researchers have shown that consistent and strong leadership is important for developing teachers (Hallinger & Heck, 1998; Heck, Larsen, Marcoulides, 1990; Reitzug, West, & Angel, 2008; Siens & Ebmeier, 1996). Therefore, it is understandable that these leadership changes negatively influenced teacher support of the new evaluation system at the middle school.

Additionally, teachers from both schools faced challenges with buy-in to the new system particularity because of concerns about the data component, with skepticism around how teachers can be held accountable for their students when there are many other factors that impact achievement (Participant L; Participant N; Participant T; Participant P; Participant I; Participant F). This point echoes aforementioned research studies that revealed legitimate concerns with including student achievement data as a component of a teacher evaluation system (Bock, Wolfe, & Fisher, 1996; Center, 2004; McCaffrey, Sass, Lockwood, & Mihaly, 2009; Toch & Rothman, 2008). Both schools tried to address these concerns by looking at pre- and post-test data around a mix of
national, state, and local assessments. This strategy is in line with previous research that point to the use of value-added assessments (Ballou, Sanders, & Wright, 2004; Hershberg, 2005; Sanders, 2000; Hershberg, Simon, & Lea-Kruger, 2004). Moreover, both schools used a growth-based model that examines data across multiple years, which reinforce previous studies that have shown more accurate estimation of teacher effectiveness by including multiple years of student achievement data (Ballou, Sanders, & Wright, 2004; Tennessee, 2004).

The other common challenges cited by both schools is meeting the amount of extra time needed in implementing the new evaluation system. Teachers particularly expressed concerns with finding time to balance their teacher duties with their additional evaluation responsibilities. On the other hand, these teachers also felt that more time should have been devoted for professional training on the new model. Addressing these two concern in the near future will be important for both schools because earlier research from the Joint Committee on Standards for Educational Evaluation (2009) established that a fair and sound evaluation system is “efficient in their use of time and resources” (p. 6). Other studies also showed the important of providing ongoing training when implementing a new evaluation model (Heneman, Milanowski, Kimball, & Odden, 2006; Sartain, Stoelinga, & Brown, 2009).

**Research Question #4: Evidence of Impact**

Both campuses did not have comprehensive data that showed direct impact of using the new evaluation model due to limited number of years of implementation. However, these schools pointed to indirect effects such as creating a school culture focused on greater teacher accountability with emphasis on data, self-reflection, and
collaboration. Leaders from both schools subsequently felt that this change of climate will positively impact student achievement in the long-term. In fact, research studies have shown such connections in these three areas. For instance, aforementioned studies show relationships between teacher accountability and student achievement (Frymier, 1998; United States Department of Education, 2002; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). Secondly, Haefele (1993) pointed out that effective teacher evaluation models hold teachers accountable while also increasing collaboration. Finally, studies have shown the positive impact of self-reflection on a teacher’s professional growth (Airasian & Gullickson, 1994; Danielson & McGreal, 2000).

At the same time, it is worth noting that this emphasis on teacher accountability and student growth also connects with the mission statements of both schools. The high school mission states: “we work together to recognize and develop potential and to promote success”, while the mission of the middle school refers to focusing on providing “academic growth” (School website). Such a connection is important as Danielson and McGreal (2000) showed that a proper evaluation system encompasses the mission and goal of the individual school and district.

**Recommendations**

After investigating the development and implementation of the teacher evaluation system at both schools, a few of my recommendations for policy makers and practitioners are given below. These recommendations are grouped into the following five categories: planning and preparation, branding and marketing of the new evaluation model, professional development, principal’s role, and differentiation.
Planning and Preparation

First, the need for planning and preparation cannot be overemphasized. Districts can avoid a number of potential issues if proper steps and procedures are set in place before implementing a new evaluation model. Such steps include creating a timeline with deadlines for the initial stages of forming and meeting with the development committee, scheduling a pilot year, and thinking through the presentation and training for both administrators and teachers on the new evaluation model. At the same time, it is critical that districts are aware of the tremendous amount of personnel and time that are necessary during these initial stages of development. Though both schools in this study were aware of this time commitment, they nonetheless still commented that they underestimated the amount of time needed to develop their new model. Such a reality further emphasizes the importance of planning and preparation.

During the initial steps of forming the evaluation committee, districts must ensure that proper representation is given to all stakeholders, especially teachers. The committee should be seen as a representation of the entire district. In fact, as the two schools in this study learned, it helps to recruit leaders and teachers who are also well-respected and influential amongst their peers because they will help provide valuable perspectives while also serving as potential future advocates of the new model.

Secondly, the committee should be aware of the time commitment necessary to accomplish the goal of developing their evaluation model. At minimum, the summer break should be allotted for committee planning and then additional time arranged for piloting the program for at least one full academic year. Piloting the program for a full academic year is necessary in order to carry out and assess the evaluation system from
beginning to end. At the same time, a number of additional teachers from different subject areas and levels should be involved in the pilot year so that they can provide a range of feedback. Such an approach also helps build an extra pool of influential supporters of the new evaluation model. With these essential steps, planning and preparing will help save valuable time and avoid potential issues in the future.

It is worth noting from this study that overturning the negative mindset of teachers about including student growth data within an evaluation system takes time. Despite two year of implementation and this study’s two schools proactively developing a new evaluation model earlier than most other districts in Michigan, their teachers still held onto negative beliefs about including student data in their evaluations. Thus, districts should plan for and be aware that teacher buy-in of the student growth component may take several years.

**Branding and Marketing of the New Evaluation Model**

The two schools in this study can attest to the importance of purposefully maintaining a positive public perception of the new evaluation model. Bolman and Deal (2008) refers to this often-forgotten but vital realm as the symbolic frame; they describe it by saying: “What is most important is not what happens but what it means” (p. 253). So my recommendation is that districts should purposely plan for and continually remain cognizant of the marketing and branding of their new evaluation model. This includes the previously mentioned step of involving representatives from the entire district when forming the evaluation committee; public perception ought to be that the new model came from the work of all stakeholders. Additionally, it is recommended that district leaders maintain a positive message when speaking about the new evaluation model with
a focus around terms such as student progress and teacher development. It is to be noted that such a consistent message is more likely to be maintained if the same district leaders, who are invested in the new model, are kept in their position. As mentioned already, the middle school leaders admitted that changes in their superintendent and principal had a negative impact on the buy-in of teachers for the new evaluation system. Districts should consequently make a commitment to avoid leadership transitions as much as possible during the initial stages of implementation, and when such changes must happen, it is prudent to hire new leaders who are invested in and acclimated with the positive messaging of the evaluation model.

Once the new evaluation model is developed, my recommendation is that the districts should purposely plan to present this model collaboratively. The unveiling presentation must not be the responsibility of the central office alone. Instead, districts can incorporate their valuable resource of teachers who participated in the evaluation committee and pilot programs. Also, symbolic leaders such as the teacher union president and community leaders can play a visual role during the model’s inauguration. At the same time, it is recommended that districts unveil the new evaluation model with the perspective that they are open to feedback and future changes as needed; in other words, opportunities for ongoing feedback are provided to all stakeholders, which will be considered by the evaluation committee. In effect, my recommendations for the branding and marketing of the new model should focus on creating a culture of collaboration for all stakeholders with flexibility for changes and a mindset built around the message of providing professional growth for teachers, resulting in student progress.
Professional Development

It is imperative that when a district makes a commitment to develop a new teacher evaluation system, they must back up this commitment with ongoing professional development and training for administrators and teachers. Both schools in this study admitted that they did not provide sufficient training. For instance, the new middle school principal was given a brief introduction to the new evaluation system and then participated in classroom observations based on the Danielson framework during the summer. The principal admitted that she gained a better understanding of the evaluation system only when she started using it during the school year. It is thus vital that districts provide adequate time for principals to practice using the evaluation model and norm classroom observations with other evaluators before the start of the school year.

Regarding teacher trainings, the central office representatives from both schools commented that they could only schedule a few trainings since other commitments had to be met within an academic calendar year. Such comments reinforced the need for districts to understand the implications of implementing a new evaluation model. After all, a new evaluation model should not be compared to another district program that may require a few initial trainings. On the contrary, a new teacher evaluation model affects the district at many levels.

For instance, the Danielson framework provides a common definition of effective teaching. Such a definition is built around four domains or realms of activity that refer to distinct aspects of teaching. When a district chooses such a framework as the foundation of their teacher evaluation model, they are also buying into the fundamental tenants around which such a model is based. Leaders and teachers should therefore have a deep
understanding of their district’s beliefs of effective teaching and the behaviors necessary to sustain good instructional practice. After all, teaching is a complex activity with various behaviors, skills, and competencies; as such, I believe it is critical that district personnel understand the frameworks around the Danielson definition of high-quality teaching. Such an understanding cannot be obtained with a couple of trainings scheduled at the beginning of the year. On the other hand, professional development should be calendared throughout the year knowing that other commitments become secondary, at least for the initial year of implementation.

At the same time, the Danielson framework with the evaluation instruments help inform professional development needs and also assist in meeting school-wide goals. For instance, when the middle school set a local goal of reducing failure rates, the fourth Danielson domain helped to monitor this goal through the use of an intervention tracker that teachers submitted for the domain of professional responsibilities. The middle school principal also analyzed data trends from her classroom observations to find areas of concern, which helped inform her future professional development needs. This type of practice was not evident at the high school. My recommendation therefore is that districts purposefully put into place ongoing data analysis of evaluation results to help inform local and district-wide professional development. As such, districts will become better informed of their teachers’ needs so that trainings are tailored and differentiated throughout the school year.

My further recommendation is that districts should ensure that their administrators and teachers are trained on how the Danielson domains can help meet their school specific goals. After all, the domains serve as a structure within which teachers can
situate their actions; schools are given the flexibility to look for certain teacher actions or the makeup of their classroom environment depending on their campus specific goals. As an example, administrators are given leeway on how to evaluate the component of managing classroom procedures, under the second Danielson domain, depending on their own school specific routines and procedures. Subsequently, the Danielson framework provides leaders and teachers an additional tool to help meet their school specific goals.

It is worth noting that these trainings focused on the Danielson framework also connect with teacher development as well. For instance, when the middle school principal provided teacher training on the questioning domain of the Danielson framework, she was also developing teacher skills in this area while at the same time building greater understanding of the Danielson evaluation model. Districts should view these types of trainings as meeting multiple professional growth goals simultaneously. It is therefore my recommendation that districts create an ongoing series of professional development focused on the different domains and evaluation instruments of the new model while also making use of data from the evaluation process to help inform teacher training and help meet school specific goals.

**Principal’s Role**

This investigation of multidimensional evaluation practices at both schools showed that the principal or administrator still plays a key role throughout the development and implementation stages. Initially, the principals at both schools served on the evaluation committee that developed the model. They then played a critical role in implementing the model by continuously monitoring their instructional program through teacher observations and feedback. Thus, even though the new evaluation system
provided a greater role for teachers to help in its implementation, the principal’s role was nonetheless significant in overseeing the entire process at their campus.

To begin with, the principals at both campuses took on the role as instructional leader, who is knowledgeable about instruction as it pertains to the Danielson framework. They had the expertise to provide appropriate feedback or recommendations when necessary. In fact, the middle and high school principals reinforced the need to continually develop their instructional expertise in the Danielson evaluation model. As evaluators of teachers, this knowledge of instruction gives them the credibility for recognizing successful practices that reflect the domains of the Danielson framework. Consequently, districts should ensure that ongoing training with opportunities to norm classroom observations are provided for principals throughout the school year. These types of trainings stand in contrast to the traditional principal preparation programs that focus on the managerial aspects of their position.

Secondly, as a part of being the instructional leader of the school, the principal will need to prioritize their time for classroom observations. At both schools, the principals made sure that they deliberately planned their calendar for classroom visits. It is important to note that the evaluation model, in fact, helps ensure that principals are devoting their time to the needs of their teachers instead of focusing on other managerial duties that take them away from their fundamental responsibilities of staff development. To help fulfill this task, my recommendation is that principals make use of their school personnel or hire an operations manager to help with responsibilities for non-instructional tasks.
In essence, the role of the principal is significant when implementing an effective teacher evaluation system. While there are numerous responsibilities for a principal such as director of personnel, site and budget manager, and parent liaison, their skills in successfully implementing the teacher evaluation system by focusing on improving teacher practice is paramount for achieving school improvement goals.

**Differentiation**

My final recommendation when implementing a new teacher evaluation model is to take advantage of providing differentiation for teachers. As Tom Peters said, “‘One size never fit all. One size fits one.’” (as cited in Littky & Grabelle, 2004, p. 74). District leaders should know that their teachers are situated at different levels of expertise and competency in regards to their professional practice. Therefore, evaluation models should be built around accommodating for these differences.

First of all, the evaluation model provides a number of data sources and instruments. For instance, the high school in this study had a variety of evaluation instruments and choices for teachers, such as the option for peer coaching, video lesson, lesson study, and action research. By giving teachers this choice, I believe they become more invested in the evaluation process since it is customized to their selection. It should be noted here that such a model built around a number of evaluation instruments should include clear policies and guidelines that are provided to leaders and teachers across multiple trainings. After all, with additional options, the potential for confusion is more likely. So I believe districts should spend time developing and piloting their guidelines so that instructions are straightforward with the use of examples, visual guides, and
flowcharts. Figures 3 and 4 that are included in the previous chapter are samples of these types of helpful illustrations.

Unlike the high school model, the middle school’s evaluation system is built around providing differentiation within the actual Danielson framework, depending upon the number of years teaching and the teacher’s evaluation status. It is worth noting that the Danielson framework offers this additional advantage in flexibility of usage. For instance, new teachers or teachers who did not score a satisfactory score in the previous year are only rated on two of the Danielson domains. In doing so, new or struggling teachers with their evaluator are able to focus on developing their primary skills of classroom management and instruction before moving onto the other Danielson domains.

Consequently, the middle and high school models make use of the above mentioned methods of differentiation that I believe can be combined to provide a more targeted and collaborative teacher evaluation system. By combining these features of the two models, a new or struggling teacher is observed using a more targeted Danielson framework while the rest of the staff are given options to choose from a number of evaluation instruments. An example of such a model is shown in Figure 5. This model also incorporates feedback loops, which are meant for ongoing systematic data analysis to help inform professional development. Figure 5 is only an example of a multidimensional teacher evaluation system that can be modified and adjusted, depending on the need and situation of each school. Furthermore, districts may initially unveil a simpler model and work towards adding evaluation instruments and offering further differentiation with time; such a growth plan may take three to five years to reflect an evaluation model similar to Figure 5.
Figure 5. Sample multidimensional evaluation model.

Ultimately, such a system provides the evaluator more time to focus on teachers with greater needs while the other teachers are given freedom to choose their evaluation instruments to help improve their practice.

Ideas for Future Research

As mentioned earlier, the purpose of this study is to investigate what strategies and processes are used by Michigan schools to build multidimensional teacher evaluation systems based on Danielson’s framework. Results from this study were analyzed to provide a number of recommendations based on the experiences from the two schools. On the other hand, as noted in an earlier chapter, one of the limitations of a case study is that its results cannot be generalized to setting outside those within the study. In other words, Yin (2003b) writes, “Case studies, like experiments, are generalizable to theoretical proposition and not to populations or universes” (p. 10). Therefore, recommendations from this study should be viewed with a filter to the specific setting of
these two schools. At the same time, this reality points to the value of conducting other similar studies at different school settings to add to the research on this topic. By offering case studies from a variety of school sites using the Danielson framework, researchers and practitioners are better able to compare these studies and find similar sites to examine.

Secondly, although the middle and high school leaders in this study acknowledged the indirect impact of their teacher evaluation system on student achievement, concrete student growth data from assessment scores were not available due to limited number of years of implementation. At the same time, there are research studies that suggest mixed data results regarding the use of the Danielson framework on student achievement (Borman & Kimball, 2005; Gallagher, 2004). It would therefore be valuable to conduct research that focuses on the direct impact of the Danielson framework on student achievement. One option is to continue this study for a few more years to investigate changes in student achievement scores across these two schools. Another option is to examine other schools using this framework across multiple years in order to measure long-term student growth data.

Besides looking at the school district level, other studies could examine how teacher preparation programs in college train their students on a multidimensional evaluation system. For instance, these programs could utilize the Danielson framework to develop teacher skills, which also builds awareness of such a framework for future evaluations. Such a practice begins with the end in mind so that future teachers know how they will be evaluated once they become educators. Future studies could therefore
investigate the extent to which teacher preparation programs prepare their students for the evaluation process.

Finally, this case study is built around the lens of looking at the process of implementing a teacher evaluation system. As Merriam (1998) wrote, “Case study is a particularly suitable design if you are interested in process” (p. 33). As such, its conclusions point to recommendations focused on the process itself. There are other angles that future case studies could examine in-depth when investigating a new teacher evaluation system. For instance, potential studies could focus on the experience of district leaders such as the superintendent or principal as they develop and implement a multidimensional teacher evaluation system. Alternatively, further studies could look at the implementation of a multidimensional evaluation system provided for administrators. After all, principals should be evaluated using a similar system as their teachers. Lastly, other studies could focus on the impact on technology during the implementation of a new evaluation model. For instance, such studies could look into the creation and use of data systems to collect and share data that is filtered to the specific needs of school leaders and teachers.

In conclusion, the results and recommendations from this study provide new insight into the implementation process of two Michigan schools using the Danielson framework as a part of their multidimensional teacher evaluation system. The study also examined the effect of these systems on principal and teacher behavior with regards to challenges the schools faced and ultimately its impact on the campus. Such a study adds to the research base so that further knowledge about effective teacher evaluation practices
is made available to encourage replications and improvements in the teacher evaluation realm of education.
REFERENCES


goals for highly qualified teachers: Estimates by state from survey data.


doi:10.1016/j.econedurev.2005.05.008


improvement. Alexandria, VA: Association for Supervision and Curriculum Development.


measures for assessing program outcomes. *Journal of Teacher Education, 57*(2),
120-138. doi:10.1177/0022487105283796


23*(1), 57-77. doi:10.3102/01623737023001057

assessment system in a large urban school district: Results of a qualitative field
doi:10.1023/A:101139203370

for teacher competencies: The use of case studies with multiple instruments for
studying teacher competencies in multicultural classes. *Teachers and Teaching:
Theory and Practice, 16*(6), 717-733. doi:10.1080/13540602.2010.517689

Denner, P. R., Miller, T. L., Newsome, J. D., & Birdsong, J. R. (2002). Generalizability
and validity of the use of a case analysis assessment to make visible the quality of
teacher candidates. *Journal of Personnel Evaluation in Education, 16*(3), 153-

student performance: A benchmarking, generalizability, and validity study of the
use of teacher work samples. *Journal of Personnel Evaluation in Education,


Foroohar, R. (2011, November 14). What ever happened to upward mobility?: Why the U.S. has become the land of less opportunity—and what we can do to revive the American Dream. *Time, 178*(19), 26-34.


Frist, Bill. (2010, March 31). Race to the top swiftly changes education dynamic. *USA


Goldhaber, D. D, & Brewer, D. J. (1996). Evaluating the effect of teacher degree level


student achievement. CREATE annual meeting, Retrieved from http://dallasisd.org


MET Project. (2012b). *Gathering feedback for teaching: Combining high quality*


Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? 
doi:10.3102/01623737026003237


Zakaria, F. (2011, November 14). When will we learn?: Education is the engine of social mobility, and it’s sputtering. Americans have two options: improve our schools or accept a lower standard of living. *Time, 178*(19), 42-44.

Appendix A

Interview Protocol
APPENDIX A
Interview Protocol

Interview

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 Western Michigan University to complete my research study and dissertation on how schools that utilize Charlotte Danielson’s research-supported Framework for Teaching are attempting to build a multidimensional teacher evaluation system that includes the new Michigan mandate to incorporate evidence of student learning as a significant element in teacher performance review systems. Therefore, the purpose of this interview is to learn more about teacher evaluation 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?

A. History & Theory of Action
1. Please briefly describe the teacher evaluation process at your school.
2. Tell me about the history of teacher evaluation at your school.
   (Possible Follow-up Questions: How/why did it get started, who were the people initially involved in developing the practice?)
3. What is the goal of teacher evaluation?
4. How is the model addressing each of Danielson’s four domains of teacher practice?
5. What other sources of information are used besides principal observations when evaluating teachers?
6. Please tell me a little about your role as (central office personnel OR a principal OR lead teacher) with respect to teacher evaluation.
B. Implementation Details
7. Who have been the main people involved with the planning and implementation of teacher evaluation practice?
8. How much start up/planning time was needed to implement your teacher evaluation model?
9. In your opinion, what factors have contributed to the successful implementation of teacher evaluation?

C. Reflection Questions
10. How do you think your teacher evaluation will lead to school improvement and higher student achievement?
11. How do you know teacher evaluation is making a difference?
12. How has the incorporation of multiple data sources into the evaluation model made a difference, positively and/or negatively?
13. What challenges did your schools face in implementing your evaluation system and how were they addressed? (Follow-up Questions: Challenges for administrators, staff, students, and parents)
14. What lessons have you learned after implementing your teacher evaluation model?

D. Resource Needed
15. What resources are needed to implement your teacher evaluation system?
16. How much professional development time has been devoted to implementing the model?
17. Do you think the training/professional development that has been conducted meets the needs for people to implement teacher evaluation? (Possible Follow-Up Questions: What other types of PD’s do you think would be helpful to effectively implement promising practice?)
18. Briefly describe to me the process that a teacher will go through when being evaluated?
19. What do you think about the use of additional data sources in the evaluation model, such as peer review, student/parent feedbacks?
20. How does the model address concerns about the legitimacy and accuracy of using student growth data when evaluating teachers?
21. What do you see as the next steps for ensuring sustainability of the evaluation model?
22. 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 B

Focus Group Protocol
[Introduction]
Thank you for taking your time to meet with me. I am working with Western Michigan University to complete my research study and dissertation on how schools that utilize Charlotte Danielson’s research-supported Framework for Teaching are attempting to build a multidimensional teacher evaluation system that includes the new Michigan mandate to incorporate evidence of student learning as a significant element in teacher performance review systems. Therefore, the purpose of this interview is to learn more about teacher evaluation 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 30 minutes. Do you have any questions for me before we begin?

Introductions:
1) To begin, please introduce yourself with your name, how long you’ve been at this school, and your position.

A. Evidence of Impact
2) What has been the impact of the current teacher evaluation system on teachers? ( “Would you explain further?” or “Is there anything else?”; “Does anyone see it differently?” or “Are there other points of views?”)
3) What benefits have you experienced as a result of implementing the current evaluation system?
(“Would you explain further?” or “Is there anything else?”; “Does anyone see it differently?” or “Are there other points of views?”)

4) What challenges have you experienced while implementing the current evaluation system?
(“Would you explain further?” or “Is there anything else?”; “Does anyone see it differently?” or “Are there other points of views?”)

5) Have there been any efforts to improve the effectiveness of teacher evaluation? If yes, explain.
(“Would you explain further?” or “Is there anything else?”; “Does anyone see it differently?” or “Are there other points of views?”)

6) What future steps are needed to ensure the sustainability of teacher evaluation?
(“Would you explain further?” or “Is there anything else?”; “Does anyone see it differently?” or “Are there other points of views?”)

7) What recommendations would you make to other educators that are thinking about adopting teacher evaluation?
(“Would you explain further?” or “Is there anything else?”; “Does anyone see it differently?” or “Are there other points of views?”)

8) Additional Comments

[Closing]
Thank you for taking your valuable time to answer these questions. Before ending our session together, have I missed anything in my questions or would you like to add one last thing?

Again, I cannot emphasize enough how your comments and answers will be invaluable for my research study. Thank you very much for your time.
Appendix C

Evaluation Implementation Observation Protocol
APPENDIX C

Evaluation Implementation Observation Protocol

School Name: ____________________________

Date: ________________________________

Topic: ________________________________

Activity Location: ____________________

Time Started: _____ Time Ended: _______ Total Time (minutes): ____

Number of Participants: _________

A. Professional Development Leadership
1. List the names and positions of participants:

B. Professional Development Session
2. Describe the intended purpose of the session.

3. List materials used for the professional development session.  
   [Note: Collect all that are available]

4. List the agenda items for the session  
5. Personal Reflections:
Appendix D

Professional Development Observation Protocol
APPENDIX D

Professional Development Observation Protocol

School Name: ____________________________
Date: __________________________________________________________________________
Professional Development Topic: ____________________________________________________
Activity Location: __________________________
Time Started: _____ Time Ended: _______ Total Time (minutes): _____
Number of Participants: ________

A. Professional Development Leadership

1. Who led training (check all that apply)?
   - __ Teacher (from the school site)
   - __ Administrator (from the school site)
   - __ Teacher from another school
   - __ Administrator from another school
   - __ University faculty member
   - __ Outside consultant (describe)
   - __ Other (describe)

List the names and positions of professional development session leaders:

B. Professional Development Session

2. Describe the intended purpose of the session.

3. Describe the content of the professional development session in detail
   (Probes: Key terms, theories, and implementation issues related to promising practice)

4. List materials used for the professional development session.
   [Note: Collect all that are available]

5. List the agenda items for the session

6. Personal Reflections:
Appendix E

Portfolio Artifact Matrix
## APPENDIX E

Portfolio Artifact Matrix

<table>
<thead>
<tr>
<th>Artifact</th>
<th>Plan/Prepare</th>
<th>Environment</th>
<th>Instruction</th>
<th>Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Individual case studies</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Examples of lesson plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Examples of unit plans reflecting appropriate content standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Documentation that demonstrates data-driven instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Documentation of the application of professional development experiences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Documentation of reflection upon peer-to-peer coaching experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Co-curricular/collaborative records, notes, materials, etc. demonstrating the artifact's involvement in the planning process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Individual student intervention plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Perception data (student &amp; parent surveys, student reflection sheets, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>10. Curriculum/class expectations (including grading policy, late work, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II. &quot;Walk-through&quot; notes and/or observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Discipline referral data, behavior charts, parent contacts, etc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Seating charts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Substitute plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Student work samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Photo documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Misc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix F

Online Survey
APPENDIX F

Online Survey

Directions: Please indicate your opinion of the following statements about the teacher evaluation system by choosing the appropriate box. The scale used includes: strongly agree (SA), agree (A), disagree (D), strongly disagree (SD), and don’t know (DK).

All surveys are to be completed anonymously and will be handled in a confidential manner.

I. Background Information
Years of experience with school_______
Total years of teaching experience________

II. Fairness
The evaluation system is a fair and efficient means for me to demonstrate my performance of professional responsibilities.
The evaluation system gives me a more prominent role in the evaluation process.
The evaluation system promotes two-way communication between my administrators and me.

III. Usefulness
The evaluation system promotes good teaching practices.
The evaluation system encourages self-reflection about my work.
The evaluation system encourages me to change instructional practices.
The evaluation system assists me in identifying areas of strength and weakness.

IV. Feasibility
The time required to review evaluations is reasonable.
The professional development and training on the evaluation system for teachers were sufficient.
The evaluators have necessary training, knowledge, and skills to conduct evaluations.
The policies regarding guidelines, purposes, and procedures of the evaluation system are clearly written and provided.
The resources needed to implement the evaluation system were sufficient.

V. Accuracy
The Danielson Framework provides an accurate representation of teacher practice.
The evaluation system is a means to provide evidence of my fulfillment of professional responsibilities not readily observable.
The evaluation system provides a richer and more comprehensive picture of my performance of the professional responsibilities.
The evaluation system includes an adequate number of data sources and inputs to provide a comprehensive view of my performance.
VI. General
To what degree did the former observation-only evaluation system contribute to your professional growth?

A great deal  Somewhat  A little  Not at all  Don’t know

To what degree do you think will the new evaluation system lead to school improvement and higher student achievement?

A great deal  Somewhat  A little  Not at all  Don’t know

To what degree did the personnel evaluations lead to appropriate professional development training?

A great deal  Somewhat  A little  Not at all  Don’t know

VII. Comments

What are the major advantages of the evaluation process from your perspective?

What are the major disadvantages of the evaluation process from your perspective?
Appendix G

Coaching Observation Form
APPENDIX G
Coaching Observation Form

<table>
<thead>
<tr>
<th>Teacher ____________________</th>
<th>Subject ____________________</th>
<th>Date __________</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coach ____________________</th>
<th>Class Period/Grade __________</th>
<th>Time-In _____</th>
<th>Time-Out ____</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 2a: Respect and Rapport    | 3a: Communicating Clearly and Accurately |
| *Teacher Interaction with students *Student Interaction | *Directions and procedures *Oral and written language |
|                            |                                              |

| 2b: Establishing a Culture for Learning | 3b: Using Questioning and Discussion Techniques |
| *Importance of content *Student pride in work *Expectations for learning and achievement | *Quality of questions *Discussion techniques *Student participation |
|                                          |                                              |

| 3b: Using Questioning and Discussion Techniques | 3c: Engaging Students in Learning |
| *Quality of questions *Discussion techniques *Student participation | *Representation of content *Activities and assignments *Groups of students *Instructional materials and resources *Structure and pacing |
|                                              |                                             |

267
<table>
<thead>
<tr>
<th>2d: Managing Student Behavior</th>
<th>3d: Providing Feedback to Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Expectations</td>
<td>*Quality: accurate, substantive, constructive and specific</td>
</tr>
<tr>
<td>*Monitoring of student behavior</td>
<td>*Timeliness</td>
</tr>
<tr>
<td>*Response to student behavior</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2e: Organizing Physical Space</th>
<th>3e: Demonstrating Flexibility and Responsiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Safety and arrangement of furniture</td>
<td>*Lesson adjustment</td>
</tr>
<tr>
<td>*Accessibility to learning and use of physical resources</td>
<td>*Response to students</td>
</tr>
<tr>
<td>*Persistence</td>
<td>*Persistence</td>
</tr>
</tbody>
</table>
Appendix H

Professional Reflection Sheet
APPENDIX H

Professional Reflection Sheet

1. Please list the focus areas that were identified in the pre-observation meeting:
   
   A. 

   B. 

   C. 

2. Please describe the lesson that was observed:

3. Please describe the strengths of the lesson:

4. Please a few areas of improvement for the lesson:

5. Please reflect on the peer coaching process. Consider including information on the following areas:
   
   a. Was the process worthwhile?

   b. What process did you use (video/direct observation/etc.)?

   c. What did you learn from the process as both an observer and an observed teacher?
Appendix I

Video Reflection Form
APPENDIX I

Video Reflection Form

Name:_________________________________ Date:____________________

Subject/Grade: ________________________ Class Period: _____________

1) After viewing the video, I observed the following behaviors as specified by the Danielson Rubric 2a-2e and 3a-3e:

2) Other effective practices I observed:

3) I observed a few areas for improvement and/or growth in the following areas:

4) Specific evidence that demonstrates whether or not:
   a) Students learned what I intended

   b) My instructional goals were met

5) Did I change my goals or instructional plan as I taught the lesson? Why or why not?

6) Based on what I observed in the video, should I do anything differently when teaching this lesson again? Why or why not?
Appendix J

Lesson Study Template
APPENDIX J
Lesson Study Template

Date:_______________________ Grade:_______

Subject___________________

Planning
Group:_________________________________________________________________________

1) Unit name
2) Unit objectives
3) Research theme (or “main aim”) of lesson study
4) Current characteristics of students
5) Lesson plan for unit:
   a. Unit goals or Outcomes (connections to Standards and Prior and subsequent learning, if appropriate)
   b. Sequence of Lessons in the Unit
   c. Explanation of unit "flow" that will enable students to move from current understanding, motivation, and skills to desired outcomes

6) Plan for the Research Lesson (The chart below may continue for several pages.)

<table>
<thead>
<tr>
<th>Teacher Activity</th>
<th>Anticipated Student Thinking and Activities</th>
<th>Points to Notice and Evaluate</th>
<th>Materials, Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   a) Aims of the lesson
   b) Learning process for the lesson (what "drama" of activities and experiences will help students move from their initial understanding to the desired aims?)
   c) Evaluation of this lesson (major points to be evaluated)
   d) Copies of lesson materials (e.g., Blackboard plan, student handouts, visual aids)

7) Background information and data collection forms for observers (e.g., seating chart, prior student work, note-taking forms, information on particular students to be observed)
Appendix K

Lesson Observation Report
## APPENDIX K

### Lesson Observation Report

**Teacher:**

**School:**

**Date of Observation:**

**Position/Subject/Period:**

**Evaluator:**

### Domain I: Planning and Preparation

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Observed</th>
<th>Not Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator 1a: Demonstrating Knowledge of Content and Pedagogy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1b: Demonstrating Knowledge of Students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1c: Selecting Instructional Goals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1d: Demonstrating Knowledge of Resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1e: Designing Coherent Instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1f: Assessing Student Learning</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

### Domain 2: The Classroom Environment

<table>
<thead>
<tr>
<th>Indicator</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator 2a: Creating an Environment of Respect and Rapport</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 2b: Establishing a Culture for Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 2c: Managing Classroom Procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 2d: Managing Student Behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 2e: Organizing Physical Space</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

### Domain 3: Instruction

<table>
<thead>
<tr>
<th>Indicator</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator 3a: Communicating Clearly and Accurately</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 3b: Using Questioning and Discussion Techniques</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 3c: Engaging Students in Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 3d: Providing Feedback to Students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 3e: Demonstrating Flexibility and Responsiveness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

### Domain 4: Professional Responsibilities

<table>
<thead>
<tr>
<th>Indicator</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator 4a: Reflecting on Teaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 4b: Maintaining Accurate Records</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 4c: Communicating with Families</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 4d: Growing and Developing Professionally</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 4e: Showing Professionalism</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
Objective:

Lesson Description:

Observed Strengths:

Recommendations for Improvement:

Other Comments/Suggestions:

Teacher Comments/Reflection attached (optional): Yes___ No___

*Teacher Signature ________________________________ Date:

Principal or Designee ________________________________ Date:
*Signature indicates only that the teacher has read this completed form.
Appendix L

Teacher Self-Checklist
APPENDIX L

Teacher Self-Checklist
(Refer to indicator elements in domain rubrics)

Teacher Name: ___________________________ Date: ___________________________

E=Exemplary  N=Needs Improvement  S=Satisfactory  U=Unsatisfactory

<table>
<thead>
<tr>
<th>Domain 1: Planning and Preparation</th>
<th>Not Observed</th>
<th>U</th>
<th>N</th>
<th>S</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator 1a: Demonstrating Knowledge of Content and Pedagogy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1b: Demonstrating Knowledge of Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1c: Selecting Instructional Goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1d: Demonstrating Knowledge of Resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1e: Designing Coherent Instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1f: Assessing Student Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments:

| Domain 2: The Classroom Environment | | | | | |
|------------------------------------| | | | | |
| Indicator 2a: Creating an Environment of Respect and Rapport | | | | | |
| Indicator 2b: Establishing a Culture for Learning | | | | | |
| Indicator 2c: Managing Classroom Procedures | | | | | |
| Indicator 2d: Managing Student Behavior | | | | | |
| Indicator 2e: Organizing Physical Space | | | | | |

Comments:

| Domain 3: Instruction | | | | | |
|-----------------------| | | | | |
| Indicator 3a: Communicating Clearly and Accurately | | | | | |
| Indicator 3b: Using Questioning and Discussion Techniques | | | | | |
| Indicator 3c: Engaging Students in Learning | | | | | |
| Indicator 3d: Providing Feedback to Students | | | | | |
| Indicator 3e: Demonstrating Flexibility and Responsiveness | | | | | |

Comments:

| Domain 4: Professional Responsibilities | | | | | |
|----------------------------------------| | | | | |
| Indicator 4a: Reflecting on Teaching | | | | | |
| Indicator 4b: Maintaining Accurate Records | | | | | |
| Indicator 4c: Communicating with Families | | | | | |
| Indicator 4d: Growing and Developing Professionally | | | | | |
| Indicator 4e: Showing Professionalism | | | | | |

Comments:

| Domain 5: Student Performance Data | | | | | |
|------------------------------------| | | | | |

Comments:

Other Comments:
Appendix M

Summative Evaluation Form: Indicators of Professional Practice
APPENDIX M

Summative Evaluation Form: Indicators of Professional Practice

Teacher Name:  
Building Assigned:  
Date of Evaluation:

E=Exemplary  
N=Needs Improvement  
S=Satisfactory  
U=Unsatisfactory

<table>
<thead>
<tr>
<th>Domain I: Planning and Preparation</th>
<th>Not Observed</th>
<th>U</th>
<th>N</th>
<th>S</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator 1a: Demonstrating Knowledge of Content and Pedagogy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Knowledge of Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Knowledge of prerequisite relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Knowledge of content-related pedagogy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1b: Demonstrating Knowledge of Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Knowledge of Characteristics of Age Group-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Knowledge of Students’ Varied Approaches to Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Knowledge of students' skills and knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Knowledge of students' interests and cultural heritage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1c: Selecting Instructional Goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Value-Goals are valuable in their level of expectations, conceptual understanding, and importance of learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Clarity-Goals are clear and permit viable methods of assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Suitability for Diverse Students-Goals are suitable for students in the class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Balance-Goals reflect several different types of learning and opportunities for integration.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator 1d: Demonstrating Knowledge of Resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Resources for Teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Resources for Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Indicator 1e: Designing Coherent Instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Learning Activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Instructional Materials and Resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Instructional Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Lesson and Unit Structure

**Indicator 1f: Assessing Student Learning**
- Instructional goals are appropriately assessed throughout the proposed approach; the approach is suitable to the goals
- Criteria and Standards

---

### Domain 2: The Classroom Environment

**Indicator 2a: Creating an Environment of Respect and Rapport**
- Teacher Interacts with Students
- Student Interaction

**Indicator 2b: Establishing a Culture for Learning**
- Importance of the Content
- Student Pride in Work
- Expectations for Learning and Achievement

**Indicator 2c: Managing Classroom Procedures**
- Management of Instructional Groups
- Management of Transitions
- Management of Materials and Supplies
- Supervision of Volunteers and Paraprofessionals

**Indicator 2d: Managing Student Behavior**
- Expectations
- Monitoring of Student Behavior
- Response to Student Behavior

**Indicator 2e: Organizing Physical Space**
- Safety and Arrangement of Furniture
- Accessibility to Learning and Use of Physical Resources

---

### Domain 3: Instruction

**Indicator 3a: Communicating Clearly and Accurately**
- Directions and Procedures
- Oral and Written Language

**Indicator 3b: Using Questioning and Discussion Techniques**
- Quality of Questions
- Discussion Techniques
- Student participation

**Indicator 3c: Engaging Students in Learning**
- Representation of Content
- Activities and Assignments
- Grouping of Students
### Domain 1: Instructional Practices

- Instructional Materials and Resources
- Structure and Pacing
- Indicator 3d: Providing Feedback to Students
  - Quality
- Indicator 3e: Demonstrating Flexibility and Responsiveness
  - Lesson Adjustment
  - Response to Students
  - Persistence

**Comments:**

### Domain 2: Assessment and Evaluation

### Domain 3: School Improvement Plan

### Domain 4: Professional Responsibilities

**Domain 4: Professional Responsibilities**

- Indicator 4a: Reflecting on Teaching
  - Accuracy
  - Use in Future teaching
- Indicator 4b: Maintaining Accurate Records
  - Student Completion of Assignments
  - Student Progress in Learning
  - Non-instructional Records
- Indicator 4c: Communicating with Families
  - Information about the Instructional Program
  - Information about Individual Students
  - Engagement of Families in the Instructional Program
- Indicator 4d: Growing and Developing Professionally
  - Enhancement of Content Knowledge and Pedagogical Skill
  - Service to the Profession
- Indicator 4e: Showing Professionalism
  - Service to Students
  - Advocacy
  - Decision Making
  - Relationships with Colleagues

**Comments:**

### Domain 5: Student Performance Data

**Domain 5: Student Performance Data**

- Constructs Appropriate Assessments
- Demonstrates Appropriate Student Achievement
- Demonstrates Measurable and Significant Academic Growth
- Uses Data to modify instruction and assessments.

**Comments:**

_____________________________________
Staff Member
(Signature signifies receipt of this final report)

___________________________________
Evaluator

___________________________________
Date: 

___________________________________
Building:
Appendix N

Danielson Rubric
**APPENDIX N**

Danielson Rubric

<table>
<thead>
<tr>
<th>Domain 1: Planning and Preparation</th>
<th>Component 1a: Demonstrating Knowledge of Content and Pedagogy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Element</strong></td>
<td><strong>Level of Performance</strong> <strong>Unsatisfactory</strong></td>
</tr>
<tr>
<td>Knowledge of Content</td>
<td>Teacher makes content errors, or does not correct content errors students make.</td>
</tr>
<tr>
<td>Knowledge of Prerequisite Relationships</td>
<td>Teacher displays little understanding of prerequisite knowledge important for student learning of the content.</td>
</tr>
<tr>
<td>Knowledge of Content-Related Pedagogy</td>
<td>Teacher displays little understanding of pedagogical issues involved in student learning of the content.</td>
</tr>
<tr>
<td>Element</td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Knowledge of Characteristics</td>
<td>Teacher displays minimal knowledge of developmental characteristics of age group.</td>
</tr>
<tr>
<td>Age Group</td>
<td></td>
</tr>
<tr>
<td>Knowledge of Students’ Varied</td>
<td>Teacher is unfamiliar with the different approaches to learning that students</td>
</tr>
<tr>
<td>Approaches to Learning</td>
<td>exhibiting, such as learning styles, modalities, and different intelligences.</td>
</tr>
<tr>
<td>Knowledge of Students’ Skills</td>
<td>Teacher displays little knowledge of students’ skills and knowledge, and does</td>
</tr>
<tr>
<td>and Knowledge</td>
<td>not indicate that such knowledge is valuable.</td>
</tr>
<tr>
<td>Knowledge of Students’ Interests</td>
<td>Teacher displays little knowledge of students’ interests or cultural heritage,</td>
</tr>
<tr>
<td>and Cultural Heritage</td>
<td>and does not indicate that such knowledge is valuable.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Element</td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td>Goals are not valuable and represent low expectations or no conceptual understanding for students. Goals do not reflect important learning.</td>
</tr>
<tr>
<td><strong>Clarity</strong></td>
<td>Goals are either not clear or are stated as student activities. Goals do not permit viable methods of assessment.</td>
</tr>
<tr>
<td><strong>Suitability for Diverse Students</strong></td>
<td>Goals are not suitable for the class.</td>
</tr>
<tr>
<td><strong>Balance</strong></td>
<td>Goals reflect only one type of learning and one discipline or strand.</td>
</tr>
</tbody>
</table>
### Domain 1: Planning and Preparation

#### Component 1d: Demonstrating Knowledge of Resources

<table>
<thead>
<tr>
<th>Element</th>
<th>Level of Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>Resources for Teaching</td>
<td>Teacher is unaware of resources available through the school or district.</td>
</tr>
<tr>
<td>Resources for Students</td>
<td>Teacher is unaware of resources available to assist students who need them.</td>
</tr>
<tr>
<td>Element</td>
<td>Un satisfactory</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Learning Activities</strong></td>
<td>Learning activities are not suitable to students or instructional goals. They do not follow an organized progression and do not reflect recent professional research. Only some of the learning activities are suitable to students or instructional goals. Progression of activities in the unit is uneven, and only some activities reflect recent professional research. Most of the learning activities are suitable to students and instructional goals. Progression of activities in the unit it fairly even, and most activities reflect recent professional research. Learning activities are highly relevant to students and instructional goals. They progress coherently, producing a unified whole and reflecting recent professional research.</td>
</tr>
<tr>
<td><strong>Instructional Materials and Resources</strong></td>
<td>Materials and resources do not support the instructional goals or engage students in meaningful learning. Some of the materials and resources support the instructional goals, and some engage students in meaningful learning. All materials and resources support the instructional goals, and most engage students in meaningful learning. All materials and resources support the instructional goals, and most engage students in meaningful learning. There is evidence of student participation in selecting or adapting materials.</td>
</tr>
<tr>
<td><strong>Instructional Groups</strong></td>
<td>Instructional groups do not support the instructional goals and offer no variety. Instructional groups are inconsistent in suitability to the instructional goals and offer minimal variety. Instructional groups are varied, as appropriate to the different instructional goals. Instructional groups are varied, as appropriate to the different instructional goals. There is evidence of student choice in selecting different patterns of instructional groups.</td>
</tr>
<tr>
<td><strong>Lesson and Unit Structure</strong></td>
<td>The lesson or unit has no clearly defined structure, or the structure is chaotic. Time allocations are unrealistic. The lesson or unit has a recognizable structure, although the structure is not uniformly maintained throughout. Most time allocations are reasonable. The lesson or unit has a clearly defined structure that activities are organized around. Time allocations are reasonable. The lesson’s or unit’s structure is clear and allows for different pathways according to student needs.</td>
</tr>
<tr>
<td>Element</td>
<td>Un satisfactory</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Learning Activities</td>
<td>Content and methods of assessment lack congruence with instructional goals.</td>
</tr>
<tr>
<td>Criteria and Standards</td>
<td>The proposed approach contains no clear criteria or standards.</td>
</tr>
<tr>
<td>Use for Planning</td>
<td>The assessment results affect planning for these students only minimally.</td>
</tr>
</tbody>
</table>
## Domain 2: The Classroom Environment

### Component 2a: Creating an Environment of Respect and Rapport

<table>
<thead>
<tr>
<th>Element</th>
<th>Un satisfactory</th>
<th>Basic</th>
<th>Proficient</th>
<th>Distinguished</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher-Student Interaction with Students</td>
<td>Teacher interaction with at least some students is negative, demeaning, sarcastic, or inappropriate to the age or culture of the students. Students exhibit disrespect for the teacher.</td>
<td>Teacher-student interactions are generally appropriate but may reflect occasional inconsistencies, favoritism, or disregard for students’ cultures. Students exhibit only minimal respect for teacher.</td>
<td>Teacher-student interactions are friendly and demonstrate general warmth, caring, and respect. Such interactions are appropriate to developmental and cultural norms. Students exhibit respect for teacher.</td>
<td>Teacher demonstrated genuine caring and respect for individual students. Students exhibit respect for teacher as an individual, beyond that for the role.</td>
</tr>
<tr>
<td>Student Interaction</td>
<td>Student interactions are characterized by conflict, sarcasm, or put-downs.</td>
<td>Students do not demonstrate negative behavior toward one another.</td>
<td>Student interactions are generally polite and respectful.</td>
<td>Students demonstrate genuine caring for one another as individuals and as students.</td>
</tr>
<tr>
<td>Element</td>
<td>Level of Performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Importance of the Content</td>
<td>Unsatisfactory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher or students convey a negative attitude toward the content, suggesting that the content is not important or is mandated by others.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher communicates importance of the work but with little conviction and only minimal apparent buy-in by the students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher conveys genuine enthusiasm for the subject, and students demonstrate consistent commitment to its value.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distinguished</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students demonstrate through their active participation, curiosity, and attention to detail that they value the content’s importance.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Pride in Work</td>
<td>Unsatisfactory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students demonstrate little or no pride in their work. They seem to be motivated by the desire to complete a task rather than do high-quality work.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students minimally accept the responsibility to “do good work” but invest little of their energy in the quality of the work.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students accept teacher insistence on work of high quality and demonstrate pride in that work.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distinguished</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students take obvious pride in their work and initiate improvements in it, for example, by revising drafts on their own initiative, helping peers, and ensuring that high-quality work is displayed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectations for Learning and Achievement</td>
<td>Unsatisfactory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructional goals and activities, interactions, and the classroom environment convey only modest expectations for student achievement.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructional goals and activities, interactions, and the classroom environment convey inconsistent expectations for student achievement.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructional goals and activities, interactions, and the classroom environment convey high expectations for student achievement.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distinguished</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both students and teacher establish and maintain through planning of learning activities, interactions, and the classroom environment high expectations for the learning of all students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Element</td>
<td>Unsatisfactory</td>
<td>Basic</td>
<td>Proficient</td>
<td>Distinguished</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Management of Instructional Groups</td>
<td>Students not working with the teacher are not productively engaged in learning.</td>
<td>Tasks for group work are partially organized, resulting in some off-task behavior when teacher is involved with one group.</td>
<td>Tasks for group work are organized, and groups are managed so most students are engaged at all times.</td>
<td>Groups working independently are productively engaged at all times, with students assuming responsibility for productivity.</td>
</tr>
<tr>
<td>Management of Transitions</td>
<td>Much time is lost during transitions.</td>
<td>Transitions are sporadically efficient, resulting in some loss of instructional time.</td>
<td>Transitions occur smoothly, with little loss of instructional time.</td>
<td>Transitions are seamless, with students assuming some responsibility for efficient operation.</td>
</tr>
<tr>
<td>Management of Materials and Supplies</td>
<td>Materials are handled inefficiently, resulting in loss of instructional time.</td>
<td>Routines for handling materials and supplies function moderately well.</td>
<td>Routines for handling materials and supplies occur smoothly, with little loss of instructional time.</td>
<td>Routines for handling materials and supplies are seamless, with students assuming some responsibility for efficient operation.</td>
</tr>
<tr>
<td>Performance of Non-instructional Duties</td>
<td>Considerable instructional time is lost in performing non-instructional duties.</td>
<td>Systems for performing non-instructional duties are fairly efficient, resulting in little loss of instructional time.</td>
<td>Efficient systems for performing non-instructional duties are in place, resulting in minimal loss of instructional time.</td>
<td>Systems for performing non-instructional duties are well established, with students assuming considerable responsibility for efficient operation.</td>
</tr>
<tr>
<td>Supervision of Volunteers and Paraprofessionals</td>
<td>Volunteers and paraprofessionals have no clearly defined duties or do nothing most of the time.</td>
<td>Volunteers and paraprofessionals are productively engaged during portions of class time but require frequent supervision.</td>
<td>Volunteers and paraprofessionals are productively and independently engaged during the entire class.</td>
<td>Volunteers and paraprofessionals make a substantive contribution to the classroom environment.</td>
</tr>
<tr>
<td>Element</td>
<td>Un satisfactory</td>
<td>Basic</td>
<td>Proficient</td>
<td>Distinguished</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Expectations</td>
<td>No standards of conduct appear to have been established, or students are confused as to what the standards are.</td>
<td>Standards of conduct appear to have been established for most situations, and most students seem to understand them.</td>
<td>Standards of conduct are clear to all students.</td>
<td>Standards of conduct are clear to all students and appear to have been developed with student participation.</td>
</tr>
<tr>
<td>Monitoring of Student Behavior</td>
<td>Student behavior is not monitored, and teacher is unaware of what students are doing.</td>
<td>Teacher is generally aware of student behavior but may miss the activities of some students.</td>
<td>Teacher is alert to student behavior at all times.</td>
<td>Monitoring by teacher is subtle and preventive. Students monitor their own behavior, correcting one another respectfully.</td>
</tr>
<tr>
<td>Response to Student Misbehavior</td>
<td>Teacher does not respond to misbehavior, or the response is inconsistent, overly repressive, or does not respect the student's dignity.</td>
<td>Teacher attempts to respond to student misbehavior but with uneven results, or no serious disruptive behavior occurs.</td>
<td>Teacher response to misbehavior is appropriate and successful and respects the student’s dignity, or student behavior is generally appropriate.</td>
<td>Teacher response to misbehavior is highly effective and sensitive to students’ individual needs, or student behavior is entirely appropriate.</td>
</tr>
</tbody>
</table>
## Domain 2: The Classroom Environment
### Component 2e: Organizing Physical Space

<table>
<thead>
<tr>
<th>Element</th>
<th>Level of Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td><strong>Safety and Arrangement of Furniture</strong></td>
<td>The classroom is unsafe, or the furniture arrangement is not suited to the lesson activities, or both.</td>
</tr>
<tr>
<td><strong>Accessibility to Learning and Use of Physical Resources</strong></td>
<td>Teacher uses physical resources poorly, or learning is not accessible to some students.</td>
</tr>
<tr>
<td>Element</td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Directions and Procedures</td>
<td>Teacher directions and procedures are confusing to students.</td>
</tr>
<tr>
<td>Oral and Written Language</td>
<td>Teacher’s spoken language is inaudible, or written language is illegible. Spoken or written language may contain many grammar and syntax errors. Vocabulary may be inappropriate, vague, or used incorrectly, leaving students confused.</td>
</tr>
<tr>
<td>Element</td>
<td>Level of Performance</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>Quality of Questions</td>
<td>Teacher’s questions are virtually all of poor quality.</td>
</tr>
<tr>
<td>Discussion Techniques</td>
<td>Interaction between teacher and students is predominantly recitation style, with teacher mediating all questions and answers.</td>
</tr>
<tr>
<td>Student Participation</td>
<td>Only a few students participate in the discussion.</td>
</tr>
<tr>
<td>Element</td>
<td>Level of Performance</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td><strong>Representation of Content</strong></td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td></td>
<td>Representation of content is inappropriate and unclear or uses poor examples and analogies.</td>
</tr>
<tr>
<td><strong>Activities and Assignments</strong></td>
<td>Activities and assignments are inappropriate for students in terms of their age of backgrounds. Students are not engaged mentally.</td>
</tr>
<tr>
<td><strong>Grouping of Students</strong></td>
<td>Instructional groups are inappropriate to the students or to the instructional goals.</td>
</tr>
<tr>
<td><strong>Instructional Materials and Resources</strong></td>
<td>Instructional materials and resources are unsuitable to the instructional goals or do not engage students mentally.</td>
</tr>
</tbody>
</table>
### Structure and Pacing

| The lesson has no clearly defined structure, or the pacing of the lesson is too slow or rushed, or both. | The lesson has a recognizable structure, although it is not uniformly maintained throughout the lesson. Pacing of the lesson is inconsistent. | The lesson has a clearly defined structure around which the activities are organized. Pacing of the lesson is consistent. | The lesson’s structure is highly coherent, allowing for reflection and closure as appropriate. Pacing of the lesson is appropriate for all students. |

### Domain 3: Instruction

#### Component 3d: Providing Feedback to Students

<table>
<thead>
<tr>
<th>Element</th>
<th>Level of Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality: Accurate, Substantive, Constructive, and Specific</strong></td>
<td><strong>Unsatisfactory</strong></td>
</tr>
<tr>
<td><strong>Timeliness</strong></td>
<td>Feedback is not provided in a timely manner.</td>
</tr>
<tr>
<td>Element</td>
<td>Unsatfactory</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lesson Adjustment</td>
<td>Teacher adheres rigidly to an instructional plan, even when a change will clearly improve a lesson.</td>
</tr>
<tr>
<td>Response to Students</td>
<td>Teacher ignores or brushes aside students’ questions or interests.</td>
</tr>
<tr>
<td>Persistence</td>
<td>When a student has difficulty learning, the teacher either gives up or blames the student or the environment for the student’s lack of success.</td>
</tr>
<tr>
<td>Element</td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Teacher does not know if a lesson was effective or achieved its goals, or profoundly misjudges the success of a lesson.</td>
</tr>
<tr>
<td>Use in Future Teaching</td>
<td>Teacher has no suggestions for how a lesson may be improved another time.</td>
</tr>
<tr>
<td>Element</td>
<td>Level of Performance</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>Student Completion of Assignments</td>
<td>Teacher’s system for maintaining information on student completion of assignments is in disarray.</td>
</tr>
<tr>
<td>Student Progress in Learning</td>
<td>Teacher has no system for maintaining information on student progress in learning, or the system is in disarray.</td>
</tr>
<tr>
<td>Non-instructional Records</td>
<td>Teacher’s records for non-instructional activities are in disarray, resulting in errors and confusion.</td>
</tr>
</tbody>
</table>
## Domain 4: Professional Responsibilities

**Component 4c: Communicating with Families**

<table>
<thead>
<tr>
<th>Element</th>
<th>Level of Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information About the Instructional Program</strong></td>
<td>Unatisfactory: Teacher provides little information about the instructional program to families.</td>
</tr>
<tr>
<td><strong>Information About Individual Students</strong></td>
<td>Teacher provides minimal information to parents and does not respond or responds insensitively to parent concerns about students.</td>
</tr>
<tr>
<td><strong>Engagement of Families in the Instructional Program</strong></td>
<td>Teacher makes no attempt to engage families in the instructional program, or such attempts are inappropriate.</td>
</tr>
<tr>
<td>Element</td>
<td>Level of Performance</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Enhancement of Content Knowledge and Pedagogical Skill</strong></td>
<td>Un satisfactory: Teacher engages in no professional development activities to enhance knowledge or skill.</td>
</tr>
<tr>
<td></td>
<td>Basic: Teacher participates in professional activities to a limited extent when they are convenient.</td>
</tr>
<tr>
<td></td>
<td>Proficient: Teacher seeks out opportunities for professional development to enhance content knowledge and pedagogical skill.</td>
</tr>
<tr>
<td></td>
<td>Distinguished: Teacher seeks out opportunities for professional development and makes a systematic attempt to conduct action research in the teacher’s classroom.</td>
</tr>
<tr>
<td><strong>Service to the Profession</strong></td>
<td>Un satisfactory: Teacher makes no effort to share knowledge with others or to assume professional responsibilities.</td>
</tr>
<tr>
<td></td>
<td>Basic: Teacher finds limited ways to contribute to the profession.</td>
</tr>
<tr>
<td></td>
<td>Proficient: Teacher participates actively in assisting other educators.</td>
</tr>
<tr>
<td></td>
<td>Distinguished: Teacher initiates important activities to contribute to the profession, such as mentoring new teachers, writing articles for publication, and making presentations.</td>
</tr>
<tr>
<td>Element</td>
<td>Un satisfactory</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Service to Students</td>
<td>Teacher is not alert to students’ needs.</td>
</tr>
<tr>
<td>Advocacy</td>
<td>Teacher contributes to school practices that result in some students being ill-served by the school.</td>
</tr>
<tr>
<td>Decision Making</td>
<td>Teacher makes decisions based on self-serving interests.</td>
</tr>
<tr>
<td>Element</td>
<td>Unsatisfactory</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Constructs</td>
<td>The teacher does not use curriculum alignment of state standards to build assessments. Assessments contain virtually no variety of questioning.</td>
</tr>
<tr>
<td>Appropriate Assessments</td>
<td>The teacher is unable to produce little or no data to support appropriate student achievement within their classroom.</td>
</tr>
<tr>
<td>Student Achievement</td>
<td>The teacher is unable to produce data to support significant academic growth of students within their classroom.</td>
</tr>
<tr>
<td>Demonstrates</td>
<td>The teacher is unable to demonstrate instances where they have used data to modify instruction and assessments.</td>
</tr>
<tr>
<td>Measurable and Significant</td>
<td></td>
</tr>
<tr>
<td>Academic Growth</td>
<td></td>
</tr>
<tr>
<td>Uses Data to</td>
<td></td>
</tr>
<tr>
<td>modify instruction and assessments</td>
<td></td>
</tr>
</tbody>
</table>
Appendix O

Individual Development Plan
APPENDIX O

Individual Development Plan

Teacher: _______________________________________ (Signature)

Administrator: _____________________________________(Signature)

Date Created: ___________________________ Date Evaluated: ___________________________

Goal One: To develop instructional delivery skills for providing student success.

Teacher Plan: (What the teacher will do.)

Administrator Support: (What the administrators will do/provide.)

Goal Two: To provide sound classroom management that is conducive to learning.

Teacher Plan: (What the teacher will do.)

Administrator Support: (What the administrators will do/provide.)
Goal Three: To meet the requirements of MCL 380.1526 regarding 15 days of professional development over the first three years of teaching. (A day is defined as an activity of 6 hours in duration.)

Teacher Plan: (What the teacher will do.)

Administrator Support: (What the administrators will do/provide.)

Goal Four: Individualized Goal(Optional upon mutual agreement.

Teacher Plan: (What the teacher will do.)

Administrator Support: (What the administrators will do/provide.)
Appendix P

HSIRB Approval Letter
Date: January 11, 2012

To: Joseph Kretovic, Principal Investigator
    George Aramath, Student Investigator

From: Amy Naugle, Ph.D., Chair

Re: Approval not needed for HSIRB 11-11-21

This letter will serve as confirmation that your project “Investigating Practices of Research-Proven Multidimensional Teacher Evaluation Systems in Michigan Schools” has been reviewed by the Human Subjects Institutional Review Board (HSIRB). Based on that review, the HSIRB has determined that approval is not required for you to conduct this project because you are studying an evaluation system and not collecting personal identifiable, private information about an individual. Thank you for your concerns about protecting the rights and welfare of human subjects.

A copy of your protocol and a copy of this letter will be maintained in the HSIRB files.