The Extent to Which Annual Professional Performance Reviews Change Classroom Instructional Practice: A Sequential Mixed-Methods Study of Teacher Evaluations in Central New York

Susan T. Vickers
St. John Fisher College, stv04448@students.sjfc.edu

Follow this and additional works at: https://fisherpub.sjfc.edu/education_etd

Part of the Education Commons

How has open access to Fisher Digital Publications benefited you?

Recommended Citation


Please note that the Recommended Citation provides general citation information and may not be appropriate for your discipline. To receive help in creating a citation based on your discipline, please visit http://libguides.sjfc.edu/citations.

This document is posted at https://fisherpub.sjfc.edu/education_etd/242 and is brought to you for free and open access by Fisher Digital Publications at St. John Fisher College. For more information, please contact fisherpub@sjfc.edu.
The Extent to Which Annual Professional Performance Reviews Change Classroom Instructional Practice: A Sequential Mixed-Methods Study of Teacher Evaluations in Central New York

Abstract
An exploratory, sequential, mixed-methods study examined the extent to which teachers of English, math, science, and social studies of Grades 9-12 change classroom practices based upon the implementation of the Annual Professional Performance Review (APPR) mandated by New York State Education Law §3012-c. Qualitative data in phase one of the study informed the survey in phase two. Critical components of classroom practice were identified in a focus group where five lead evaluators organized the study into two themes: lesson planning and classroom atmosphere. A second focus group of six teachers identified 20 impetuses that promote changes in classroom instructional practice. The survey received 292 responses from teachers, of which 86 met the criteria to fully participate. Lesson planning was identified as the area where teachers made the most changes during the 2013-2014 and 2014-2015 school years. Collaboration with peers within their own district was identified as the primary impetus by teachers to change practice. APPR was ninth on the list of impetuses for changing practice. Understanding what drives teachers to change practice provides districts with the knowledge necessary for guiding teachers to fully develop classroom instructional strategies.

Document Type
Dissertation

Degree Name
Doctor of Education (EdD)

Department
Executive Leadership

First Supervisor
C. Michael Robinson

Second Supervisor
Idonia Owens

Subject Categories
Education

This dissertation is available at Fisher Digital Publications: https://fisherpub.sjfc.edu/education_etd/242
The Extent to Which Annual Professional Performance Reviews Change Classroom Instructional Practice: A Sequential Mixed-Methods Study of Teacher Evaluations in Central New York

By

Susan T. Vickers

Submitted in partial fulfillment of the requirements for the degree Ed.D. in Executive Leadership

Supervised by

C. Michael Robinson, Ed.D.

Committee Member

Idonia Owens, Ed.D.

Ralph C. Wilson, Jr. School of Education

St. John Fisher College

December 2015
Dedication

Successful completion of my doctoral degree is testament to the core values instilled in me by my father, Robert D. Vickers. Thus, this study is dedicated posthumously to my dad who sacrificed much in order to provide me with a well-rounded and rich childhood. My father believed in the value of an education, as a solid education opens doors to limitless opportunities. As an avid reader, he was a role model for those who are termed “life-long learners.” I remember being in awe with the depth of his knowledge on a wide array of topics, especially regarding military history. Always my father’s daughter, I developed a passion for history and earned a Bachelor’s and Master’s Degree in Military History.

My father was well schooled in mastering his craft, be it at work, raising a Rottweiler, or tending to his garden. This ability to focus on a topic to ensure success is the root of my interest in earning a doctoral degree in executive leadership. Earning the degree is just the beginning of the journey. As to the next step and what will be done with the degree is yet to be determined. This step is the natural progression of my father’s questioning. Should I have been so fortunate to have my father at my doctoral graduation, he would have asked, “So what now?” Indeed, what now?
Biographical Sketch

Susan T. Vickers is currently the Associate Principal of Cazenovia High School and Director of Secondary Curriculum for Cazenovia Central School District. Miss Vickers attended the State University of New York at Brockport, graduating with a Bachelor of Science in History and Political Science in 1986. She returned to Brockport in 1992 and earned a Master of the Arts degree in History. In 2004, Miss Vickers earned a Certificate of Advanced Study in Education Administration from the State University of New York at Cortland. Her doctoral studies began in the fall of 2013 at the Onondaga Community College extension site of St. John Fisher College Ed.D. Program in Executive Leadership. Miss Vickers pursued her research in New York State’s Annual Professional Performance Review Program under the direction of Dr. C. Michael Robinson and received her Ed.D. Degree in 2015.
Abstract

An exploratory, sequential, mixed-methods study examined the extent to which teachers of English, math, science, and social studies of Grades 9-12 change classroom practices based upon the implementation of the Annual Professional Performance Review (APPR) mandated by New York State Education Law §3012-c. Qualitative data in phase one of the study informed the survey in phase two. Critical components of classroom practice were identified in a focus group where five lead evaluators organized the study into two themes: lesson planning and classroom atmosphere. A second focus group of six teachers identified 20 impetuses that promote changes in classroom instructional practice. The survey received 292 responses from teachers, of which 86 met the criteria to fully participate. Lesson planning was identified as the area where teachers made the most changes during the 2013-2014 and 2014-2015 school years. Collaboration with peers within their own district was identified as the primary impetus by teachers to change practice. APPR was ninth on the list of impetuses for changing practice. Understanding what drives teachers to change practice provides districts with the knowledge necessary for guiding teachers to fully develop classroom instructional strategies.
Table of Contents

Dedication ........................................................................................................................................ iii
Biographical Sketch ...................................................................................................................... iv
Abstract .......................................................................................................................................... v
Table of Contents ......................................................................................................................... vi
List of Tables ................................................................................................................................... ix
List of Figures ............................................................................................................................. x

Chapter 1: Introduction ................................................................................................................ 1
  Introduction ................................................................................................................................. 1
  Problem Statement ................................................................................................................... 2
  Theoretical Rationale ............................................................................................................. 4
  Statement of Purpose .............................................................................................................. 5
  Research Questions .................................................................................................................. 6
  Potential Significance of the Study ....................................................................................... 7
  Definitions of Terms ............................................................................................................... 8
  Chapter Summary .................................................................................................................... 8

Chapter 2: Review of the Literature ........................................................................................... 10
  Introduction and Purpose ....................................................................................................... 10
  Review of the Literature ....................................................................................................... 10
  Chapter Summary ................................................................................................................... 63

Chapter 3: Research Design Methodology ................................................................................ 65
### List of Tables

<table>
<thead>
<tr>
<th>Item</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Permissible Statewide Ranges for Teacher Evaluations</td>
<td>43</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>State Prescribed Conversion Table Assigning Student Growth Scores</td>
<td>44</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>The Matrix</td>
<td>45</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>New York State Public School RTTT Allocations</td>
<td>50</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Public School Districts Within the Central New York Region</td>
<td>70</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>105Changes Reported in Classroom Instructional Practices</td>
<td>105</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Top Six Impetuses For Changing Lesson Planning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error! Bookmark not defined.</td>
<td></td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Impetuses for Changing Classroom Instructional Practices in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lesson Planning</td>
<td>109</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Top Five Impetuses For Changing Classroom Atmosphere</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error! Bookmark not defined.</td>
<td></td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Impetuses for Changing Classroom Instructional Practices in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classroom Atmosphere</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error! Bookmark not defined.</td>
<td></td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Item</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>The Contextual Factors, Constructs, and Measures Associated with Teaching Quality</td>
<td>31</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Exploratory Sequential Mixed-Methods Flow Chart</td>
<td>67</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Thematic Coding Identification of Lesson Planning as Identified by Five Lead Evaluators on July 27 2015</td>
<td>91</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Thematic Coding Identification of Classroom Atmosphere and Experience as Identified by Five Lead Evaluators on July 27, 2015</td>
<td>94</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Data Identified from Coding Teacher Focus Group Discussion on What has Caused Them to Change Classroom Instruction Over the Past 10 Years</td>
<td>97</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>The Study Included an Even Distribution of Participation From the Four Cour Subject Areas</td>
<td>103</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Degree to Which Participants Changed Classroom Instructional Practice Based Upon the List of Identified Impetuses</td>
<td>106</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Degree to Which Participants Made Changes to the Classroom Instructional Practice, the Classroom Atmosphere, Based Upon the List of Identified Impetuses</td>
<td>110</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Introduction

Public education has been at the forefront of national and state debate for decades as student academic achievement is not on par with other nations. Policies holding teachers accountable for student achievement include implementing rigorous teacher evaluation systems (Bereens, 2000; Ravitch, 2010; Staiger & Rockoff, 2010). Ravitch (2010) stated that accountability is the nation’s education policy. In order for school improvement to succeed, schools need effective teachers. Bereens (2000) offered that teacher evaluation is one of the most complex and important issues in education. Performance evaluations are the primary tool school districts employ to improve classroom instruction. Teacher evaluations focus on teachers’ learning, as it supports teachers in professional reflection and development of their instructional practices (Benedict, Thomas, Kimerling, & Leko, 2013).

Including student achievement data into overall performance evaluations hold teachers accountable for student learning. “Teacher evaluations are powerful tools that can inhibit or empower employees in relation to quality and productivity” (Law, 2007, p. 19). Therefore, it is imperative for evaluation tools to do more than identify strong and weak teachers (Papay, 2012). The New York State Annual Professional Performance Review (APPR) plan includes providing an overall rating for teachers on a scale denoting their status as highly effective, effective, developing, ineffective (EngageNY, 2012a).
New York State Education Law §3012-c requires an evaluation of all public school teachers under the newly revised NYS APPR process. Consequently, districts are able to share $696,646,000 from federal grant money as a result of the state implementing rigorous evaluation systems (New York State Education Department [NYSED], 2010). There is a moral responsibility for state education and elected state representatives to ensure that the expenditure of these tax dollars are reaping the intended benefits. To be effective, teacher evaluation tools must assist teachers in assessing their performance as well as provide a road map for developing pedagogical skills that will improve achievement for each student.

**Problem Statement**

Effective teachers have the power to transform learning for every student (Tucker & Stronge, 2005). School principals evaluate teachers to improve classroom practices and to guide overall professional development. Accountability for student learning is ascertained through teacher evaluations. Kersten and Israel (2005) contend that the national focus on accountability in education occurred with the publication of *A Nation at Risk* in 1983. Accountability remains a national focus with continued reauthorization of the Elementary and Secondary Education Act (ESEA) (2009) by Congress.

With congressional reauthorization of ESEA in 2009, President Barack Obama established a goal to have a great teacher in every classroom and a great principal in every school (U.S. Department of Education [USDE], 2010). The Race to the Top (RTTT) Fund (2009) was subsequently established to help achieve this goal by providing competitive grants to states that enacted ambitious and comprehensive reforms that focused specific attention on teacher accountability.
New York State Education Law §3012-c addresses teacher evaluation and accountability in order to qualify for RTTT grant money. By September 1, 2011, public school districts across New York State adopted APPR plans for instructional staff in compliance with the law (EngageNY, 2012b). Effective in the 2012-2013 school year, all instructional teachers were included in the APPR process. Individual districts and the New York State Education Department (NYSED) shared $696,646,000 in federal grant money from the RTTT Fund (NYSED, 2010). This grant money was in addition to annual state aid that was provided to districts across New York State.

Performance appraisals are powerful tools that can inhibit or empower employees in relation to quality and productivity (Law, 2007). Teacher evaluation is a complex issue and an important issue in education that can be frustrating for both teachers and administrators (Bereens, 2000). The complexity is magnified when evaluators try to quantify instructional nuances that are often difficult to observe through normal evaluation procedures. Teachers are the most critical factor in regard to student learning, and quantitative research illustrates the wide variation in teachers’ abilities to impact student test scores (Rockoff, 2004).

It is imperative for evaluation tools to do more than identify strong and weak teachers (Papay, 2012). To be effective, teacher evaluation tools must assist teachers in assessing their performance as well as provide a road map for how to develop pedagogical skills that will improve student achievement. Teacher evaluations, theoretically, cause teachers to change their classroom instruction and drive professional development. Bereens (2000) stated that the ability of teachers to trust their leaders predicates any systemic or effective change in their classroom practice.
Research supports the notion that the instrumental factor in student success is the classroom teacher. Federal and state legislative attempts to address the issue focus on holding teachers accountable for student success. While research supports teacher evaluations as a formative tool to improve classroom performance on the part of the teacher, there is little empirical evidence regarding the impact of APPR (3012-d) on changes in teachers’ classroom instructional practices to improve student learning.

The purpose of this study is to examine the extent to which public school teachers of social studies, science, English, and math in Grades 9-12 are changing classroom practices based upon the implementation of New York State Education Law §3012-c, known more commonly as APPR. Considerable amounts of time, effort, and financial expenditures on the part of school districts are spent implementing APPR in the hopes of changing teachers’ practice to improve student learning. Therefore, it is critical to understand the extent to which APPR is changing instructional practice in the classroom.

**Theoretical Rationale**

Performance appraisals provide adults with an opportunity to receive feedback that may assist in improving their productivity (Ponticell & Zepeda, 2004). Ponticell and Zepeda (2004) stated that adults are performance-centered, and this aspect drives their desire for continued improvement. Constructive feedback is a key component of the newly implemented APPR system in New York State.

Performance feedback theory holds that individuals and organizations cyclically evaluate and compare their performance to previous history or comparable models as a form of continuous improvement (Greve, 2003). Based upon this cyclical review,
decisions are made to either change practice or continue with the current practice. The goal of the individual or organization is to attain and maintain performance expectations.

Performance feedback theory provides the framework to review teacher evaluations. Annual evaluations of teachers occur with the goal of improving performance in the classroom. This theory shares basic elements that are included in teacher evaluation rubrics: (a) goal-setting, (b) performance feedback, and (c) establishment of aspiration levels to change practice (Greve, 2003).

Performance feedback theory is a diagnostic tool used to discover problems (Ryan, 2004), and it is a cyclical process for continuous improvement (Greve, 2003). APPR employs the same tenets of performance feedback theory with regard to identifying areas of concern for classroom teachers. APPR provides school districts with information that is designed to drive professional development for faculty and staff. Focusing on continuous improvement through feedback is synonymous with APPR and performance feedback theory. According to Greve (2003), when goals and aspiration levels are attained, decision makers are less inclined to take action to improve performance. Similarly, APPR is designed to address complacency in schools.

**Statement of Purpose**

The purpose of this study is to examine the extent to which public school teachers of social studies, science, English, and math in Grades 9-12 are changing classroom practice based upon implementation of New York State Education Law §3012-c. Considerable amounts of time, effort, and financial expenditure on the part of school districts are spent implementing APPR in the hopes of changing teacher practice to
improve student learning. Therefore, it is critical to understand the extent to which APPR is changing instructional practice in the classroom.

**Research Questions**

This study examines the extent to which public school teachers of English, math, science, and social studies in Grades 9-12 are changing classroom practice based upon implementation of NYS Education Law §3012-c, known more commonly as the APPR. Collected data answers the following questions.

1. From the perspective of selected, experienced lead evaluators within Central New York school districts, who have 2 years of experience under the new APPR process, what are the critical components of classroom instructional practice contained in the state-approved evaluation rubrics?
2. From the perspective of selected teachers, in Grades 9-12 within Central New York school districts, who have 2 years of experience under the new APPR process, what are the primary impetuses for teachers to change classroom instructional practice?
3. Given the critical components of classroom practice, as identified by question 1, from the perspective of selected English, math, science, and social studies teachers in Grades 9-12 within Central New York school districts, in which component(s) did they make the most changes in instructional practice during the 2013-2014 and 2014-2015 school years?
4. Given the changes identified by the same participants in question 3, what was the primary impetus for changes made in their classroom instructional practice?
Potential Significance of the Study

Effective teachers have the power to transform learning for every student (Tucker & Stronge, 2005). To become an effective teacher requires purposeful work on the part of teachers and principals who work together as reflective practitioners. Principals observe teachers as part of a performance evaluation process designed to improve classroom practices and to guide overall professional development for the school district. The power of feedback to improve instruction is influenced by the instruments and procedures designed for the evaluation process (Hill & Grossman, 2013).

In order to receive Race to the Top funding for education, states have to implement teacher evaluation systems that replaced dichotomous satisfactory and unsatisfactory teacher ratings with multiple-category ratings for teachers (Darling-Hammond, 2013). Governor Andrew Cuomo of New York State accepted the federal challenge to implement an evaluation system that held teachers and principals more accountable for student learning and passed NYS Education Law §3012-c on May 28, 2010 (NYSED, 2015). In turn, the New York State Education Department was awarded $696,646,000 in federal grant money from RTTT (NYSED, 2010). By September 1, 2011, public school districts across New York State had to adopt APPR plans for all instructional staff (EngageNY, 2012b).

The significance of the study focuses on the overall affect that the APPR program has in regard to improving student learning by holding teachers more accountable. “With a graduation rate of 76% in 2010, New York State implemented the APPR process with a goal of improving student learning (NYSED, 2011).” Lessons learned from the study
focus on what drives instructional changes in the classroom along with the overall impact of the APPR process on teacher’s instructional practice.

Definitions of Terms

*Lead Evaluator* – those individuals trained to assess teachers who are using state approved rubrics (OCM BOCES, 2011). Danielson (2010b) interchangeably uses the term *skilled evaluator*. Evaluators “must be able to recognize classroom examples of the different components of practice, interpret that evidence against specific levels of performance, and engage teachers in productive conversation about their practice” (Danielson, 2010, p. 36).

*Rubric* – is a scoring tool that identifies the criteria utilized in assessing a piece of work (Archer & Hughes, 2011). It is a tool that allows for assessing performance-based work by clearly denoting quality gradation.

*Standards* – includes a clear depiction of what is to be accomplished. Rosenholtz (1999) added that standards offer a “minimum achievement level used as a reference point when judging the quality of work” (p. 12).

Chapter Summary

The study focuses on the APPR process of teacher evaluations for improving student achievement as evident in the overall student graduation rate and performance on state assessments. As a formative tool, evaluations have the ability to transform teachers, and consequently improve student achievement. Assessing teachers is only part of the journey of the teacher evaluation process, as it can be used as a road map for the development of critical instructional practices to improve student achievement needs, which are to be embedded in the tools and procedures of the process. If evaluations only
include a categorization of teacher effectiveness, then they have failed in the endeavor with a significant cost to taxpayers.

Chapter 2 provides a selective review of the literature germane to performance evaluations both within the public and private sector. Included in the review of the evaluations are the critical facets and inherent biases of evaluation tools. Supporting the performance evaluation concept is a review of Heinrich Greve’s (2003) performance feedback theory.

Chapter 3 outlines an exploratory sequential mixed-methods procedure of the study. The study includes two phases, and the participants include lead evaluators and teachers of English, math, science, and social studies of Grades 9-12 from Central New York.

Chapter 4 presents the results of the sequential mixed-methods study. Data analysis for phase one of the study employed coding and descriptive statistics was utilized for phase two. Findings include a discussion of the focus groups from phase one and the quantitative survey from phase two.

Chapter 5 exhibits a discussion of the extent to which APPR promotes change in classroom instructional practice for English, math, science and social studies teachers of grades 9-12. The chapter includes discussion on the implications of findings, limitations of the study, and recommendations for future research.
Chapter 2: Review of the Literature

Introduction and Purpose

This chapter provides a review of the literature relevant to performance appraisals and, more specifically, evaluations of teachers. Concise and insightful analysis of empirical studies support key concepts and themes as they relate to the dissertation topic.

The review of literature commences with an overview of the state of educational accountability in the United States and then narrows to focus on the New York State annual professional performance review program. Evaluation of teacher performance is reviewed with attention paid to bias and administrator impact on the overall process. Presented as a supporting framework for the study is Heinrich Greve’s (2003) performance feedback theory. Of particular interest is the application of change of individual practice, which is offered in the theory, as it applies to change of classroom instructional practice resulting from teacher evaluations.

Review of the Literature

State of educational accountability. President Lyndon B. Johnson acknowledged a need for federal intervention in education with the implementation of the Elementary and Secondary Education Act (ESEA) in 1965. As part of the War on Poverty, President Johnson’s enactment of the ESEA established high standards and accountability for public education (USDE, 2013). Originally authorized into law in 1965, the ESEA has received congressional reauthorization since enactment.
Teacher evaluations have received scholarly attention for decades. Kersten and Israel (2005) stated that the call for school improvement began with the publishing of *A Nation at Risk* in 1983. Secretary of Education, Terrell H. Bell, commissioned the report in 1981 (Geier, 2014). In 1989, President George H. W. Bush furthered the nation’s attention on educational accountability when he commenced work on the reauthorization of ESEA. This culminated in The Goals 2000: Educate America Act that President Clinton later signed into law in 1994 (Geier, 2014). Goals 2000 focused on outcome-based education and included an abundance of school improvement goals of which teacher accountability was one. Since the 1990s, researchers have identified the need for standards-based, teacher-observation systems in lieu of checklists and narratives (Kersten & Israel, 2005).

President George W. Bush addressed the concern for best practice and research-based education by the reauthorization of ESEA in 2002 and the passage of No Child Left Behind Act (NCLB) (USDE, 2013). NCLB targeted four main goals: (a) accountability for results; (b) an emphasis on doing what works, based upon best practice and research; (c) expanded parental options, and; (d) expanded local control and flexibility (Geier, 2014).

Ravitch (2010) stated that “NCLB made accountability the nation’s education policy” (p. 20), and caused controversy as three states subsequently sued the United States Department of Education, arguing that the unfunded mandates violated their abilities to enact them (Geier, 2014). While the judge ruled, on appeal, that only funded mandates require compliance, the lofty tenets of NCLB remained in place as public school districts struggled to implement the law (Geier, 2014). Accountability for results
included testing students in math and English in Grades 3-8 and in science for students in Grades 4 and 8.

Congress reauthorized ESEA with the implementation of the American Recovery and Reinvestment Act (ARRA) in 2009. President Obama stated that his goal was to have a great teacher in every classroom and a great principal in every school (USDE, 2010). To assist in the implementation of ARRA, a corollary program known as Race to the Top (RTTT) (2010) was established. RTTT provided competitive grants to school districts and states that took on ambitious and comprehensive reforms. RTTT focuses specific attention on teacher accountability as it relates to student achievement. Satisfactory and unsatisfactory evaluation ratings are replaced with highly effective, effective, developing, ineffective (HEDI) categories of teacher ratings based upon multiple observations, feedback, and use of student test scores (Darling-Hammond, 2013).

The United States Government Accountability Office (GAO) reviewed progress on the $4 billion awarded to 12 states for the RTTT initiative (Government Accountability Office (GAO), 2013). Six of 12 states had fully implemented their evaluation systems, and only three met the targeted date of implementation. States reported challenges with the implementation of teacher accountability practices, evaluation measures, and building sustainability.

Since 2009, school districts across the nation have passed teacher evaluation reforms to improve their candidacy for federal funding (Laine & Behrstock-Sherratt, 2014). Increased accountability is the common denominator for school reforms in the United States (Cochran-Smith, Piazza, & Power, 2013; Ravitch, 2012). Laine and
Behrstock-Sherratt (2014) stated that as districts reform teacher evaluation and accountability, they are focusing on measures of teacher impact on student learning aside from testing.

Ravitch (2012) continued that student test scores should not be utilized to measure a teacher’s impact on student learning, as the test scores are subject to statistical errors. Ravitch contended that student test scores assess student performance on the exam, not the teacher’s performance in the classroom. RTTT conversely mandates that “student test scores count for a significant part of teachers’ evaluations, determining their tenure, promotion, and job security” (Ravitch, 2012, p. 12). According to the GAO, 11 out of the 12 RTTT states found teachers concerned about the use of student academic growth data in evaluations and the use of evaluation results to make personnel decisions (GAO, 2013). Darling-Hammond (2013) supported this concern and added that individualistic and competitive approaches that sort and rank teachers undermine the citizenship spirit of educators. By sorting and ranking teachers, a competitive environment may emerge.

Ravitch (2012) questioned the applicability of student test scores as indicative of school improvement. Mausethagen (2013) found that focusing on testing and student performance has a negative impact on the teacher-student relationship.

Mausethagen (2013) conducted a review of 28 international quantitative and qualitative studies that focused on the influence of accountability policies and increased testing on teachers’ work and teachers’ interactions with students. The selection of the studies included: (a) only articles published between 1990 and 2010, (b) peer-reviewed and published literature, (c) empirical articles in which teachers were informants, and (d) articles that included accountability and teacher role. The studies reviewed were mostly
qualitative, although six were mixed methods. As the issue of accountability is universal, the review included 10 studies from the United States, nine from the United Kingdom, three from Nordic nations, and one, each, from China, Portugal, Canada, Australia, and New Zealand. The purpose of the study was to examine the career phases of teachers.

Mausethagen’s (2013) findings stated that increased accountability contradicts the traditional view of education as being one that nurtures efficacy and student learning. The motivation to not compromise on the moral, relational, and caring aspects of being a teacher resulted in resistance to comply fully with accountability procedures.

Elstad, Knut-Andreas, and Turmo (2012) conducted a qualitative study that explored the link between the strength of accountability and teachers’ organizational citizenship behavior. The study acknowledged the role of trust as an indicator of the strength of organizational citizenship behavior (OCB). OCBs are thought of as discretionary behaviors, which are not part of the job description, and are performed by the employee as a result of personal choice. Second, OCBs go above and beyond that which is an enforceable requirement of the job description. Finally, OCBs contribute positively to overall organizational effectiveness.

In 2009 and 2010, 11 schools participated in the study by completing surveys (Elstad et al., 2012). The study identified that validity problems are related to assessments. Compounding the issue was that the primary productivity of a teacher is not readily measureable. By primary productivity, Elstad et al. (2012) were referring to a teacher as a caring individual who builds social and emotional bonds with students. The findings illustrated the need for good relations between educators and leaders (Darling-Hammond, 2013; Elstad et al., 2012). Relationships within the area of accountability are
an element of education that requires attention (Elstad et al., 2012; Mausethagen, 2013). Implementation of a rigorous new teacher appraisal program in New York State mirrors accountability and measurability concerns, which were noted by the researchers.

At the national level, officials in three states experienced difficulty conveying to teachers that the use of the new accountability procedures was to drive professional development rather than utilization for negative consequences (GAO, 2013). This statement followed a previous entry in the GAO report that noted state and district officials experienced difficulty ensuring consistency by principals as they assessed teachers’ professional practice.

It is natural to fear being evaluated (Bechar & Mero-Jaffe, 2013), which was identified by the fear of negative consequences encountered by officials as they implemented the new accountability procedures for evaluation (GAO, 2013). This fear is rationalized given that appraisal systems are only effective based upon employee reactions to the process and the outcomes they receive (Swiercz, Bryan, Eagle, Bizzotto, & Renn, 2012). Swiercz et al. (2012) identified trust and fairness as key indicators that influence employee attitudes toward the organization and appraisal system. When employees believe the process is fair, they are more receptive to the outcomes. Performance appraisals are powerful tools that can inhibit or empower employees in relation to quality and productivity (Law, 2007). Teacher evaluation is a complex and important issue in education that can be frustrating for both teachers and administrators (Bereens, 2000).

The complexity of teacher evaluations is evident in the complexity of the process for both teachers and administrators. The lack of clear and consistent standards of good
practice or focus on improving practice are part of the current problems and complexity with teacher evaluation systems (Darling-Hammond, 2013). In addition, Darling-Hammond (2013) identified that inadequate time and availability of staff to complete effective evaluations compounds issues concerned with improving teacher quality. When evaluations are not prioritized, detachment affects professional development. Darling-Hammond (2013) noted that teachers want robust evaluation systems that are fair, useful, and aligned with professional development. “Teacher’s ongoing learning, in turn, depends on the construction of a strong professional development system and useful career development approaches that can help spread expertise” (Darling-Hammond, 2013, p. 3).

Evaluating teacher performance. Teachers are the most critical factor in regard to student learning, and quantitative research illustrates the wide variation in teachers’ abilities to impact student test scores (Rockoff, 2004). Danielson (2010) offered, “A principal or superintendent must be able to say to the school board and the public: Everyone who teaches here is good – and here’s how I know” (p. 36). Succinctly answering this question provides credibility to that district’s evaluation system. There are two reasons why we evaluate teachers: (a) to ensure teacher quality, and (b) to promote professional development (Danielson, 2010; Papay, 2012).

Teacher evaluation results in professional conversations about teacher practice (Danielson, 2010; Papay, 2012). Feedback provided during post-observation meetings is one of the most critical components to foster and improve teacher performance (Danielson, 2010). Papay (2012) referred to teacher evaluation as a formative
professional development tool that identifies teachers’ strengths and weaknesses and highlights areas for improvement.

Papay (2012) reframed teacher evaluation reforms by analyzing the two major approaches to evaluation reform: value-added and standards-based evaluations. Seeking to broaden the scope of assessing teachers, two approaches were analyzed regarding their ability to inform and support ongoing teacher development.

It is imperative for evaluation tools to do more than identify strong and weak teachers (Papay, 2012). To be effective, teacher evaluation tools must assist teachers in assessing their performance as well as provide a road map for how to develop pedagogical skills that will improve student achievement. Papay (2012) viewed teacher evaluation as a professional development tool that focused on specific and meaningful feedback. This view was in contrast to the traditional view that serves to identify the best and worst teachers. Traditional views on evaluation are fraught with serious concerns of bias, reliability, and validity.

Taylor and Tyler (2011) studied whether teacher evaluation can improve teacher practice, and they were concerned with how new information from being evaluated might lead to change in teacher practice. Participants in the quasi-experimental study from the 2002-2003 through the 2009-2010 school years were mid-career elementary and middle school math teachers from the Cincinnati Public School system. Their study compared student achievement of individual teachers before, during, and after each teacher’s evaluation year.

The findings suggest that teachers can and do improve with specific and meaningful feedback (Taylor & Tyler, 2011). Students taught by a teacher during an
evaluation year scored, on average 11% of a standard deviation higher than peers taught by the same teacher in years preceding the evaluation. Taylor and Tyler (2011) stated that teachers who receive critical feedback improved student achievement scores during their evaluation year and for years after the evaluation. They postulated that the increase in student achievement for teachers who had the most room for improvement gave credence to the relationship between the evaluation process and the teacher change. Danielson (2010) also stated that feedback from the evaluation process drives the change in teacher performance.

Taylor and Tyler’s (2011) study revealed that the greatest change in student improvement occurs during the teacher’s evaluation year. They support Danielson’s (2010) premise regarding the impact of teacher evaluation. Along with Danielson (2010), they also identified the problem districts endure regarding the challenge to engage in evaluation with integrity and fidelity.

The cost to the Cincinnati School District was $1.8 million to $2.1 million per year during the years in which the evaluation study was conducted (Taylor & Tyler, 2011). The cost per teacher for the evaluation process was $7,500, with more than 90% of the funds allocated to the cost of evaluators’ salaries. “If done well, performance evaluation can be an effective form of teacher professional development” (Taylor & Tyler, para. 32). Done well, by Danielson’s (2010) standards, requires carefully setting priorities for instruction and judicious scheduling of observations and post-observation conferences (Danielson, 2010). Danielson (2010) stated that the challenge for administrators is finding time to conduct meaningful observations and for engagement in professional conversations about teacher practice.
Kimball and Milanowski (2009) conducted a sequential mixed-methods research study that focused on 23 school leaders in a western U.S. public school district with 3,300 teachers. The district employed Danielson’s (2010) standards-based evaluation system for 3 years prior to the study. Their study commenced with a statistical analysis to determine if evaluators differed in regard to evaluation ratings. Administrators who were included in this analysis subsequently became the study focus of the evaluation procedure. The researchers employed a sequential mixed-methods design to determine if the evaluators in the district differed in validity of their evaluation ratings.

Kimball and Milanowski (2009), and later supported by Darling-Hammond (2013), noted that standards-based rubrics are only as effective as the manner in which they are used. They contend that motivation, knowledge of the rubrics and instruction, and conditions within the educational community interact in complex ways that are idiosyncratic across evaluators (Kimball & Milanowski, 2009). Their 3-year longitudinal study found that teacher quality ratings varied considerably between evaluators. Subjectivity and bias on the part of the evaluator may account for varied evaluation scores (Bell, Gitomer, McCaffrey, Hamre, Pianta, & Qi, 2012; Hill & Grossman, 2013).

Feedback and assessment information from observations must be correct in order to improve performance (Tyler, 2012). The relationship between teacher quality and student assignment finds higher performing students assigned to highly effective teachers (Adamson & Darling-Hammond, 2012; Borman & Kimball, 2005). According to Borman and Kimball (2005), scoring of teacher quality appears related to what courses and students are assigned to the teacher. They further contend that perception of teaching quality may find bias in how teacher performance is ultimately scored.
Murphy, Hallinger, and Heck (2013) confronted the issue of teacher evaluation by questioning the defensibility of a score when it appears as though there is not buy-in by the evaluators. Questionable buy-in was noted with evaluators in a bad mood, evaluating courses where they either have no knowledge or expertise in, time of day that may conflict with when the evaluator functions at their best, and possibly the overall health of the evaluator. Murphy et al. (2013) further questioned teachers’ ability to trust the evaluation as a gauge of their teaching quality. Feedback from evaluations serve as the impetus for teachers to change their instructional practice (Bereens, 2000). Researchers identified trust as a critical component for sustained instructional change (Bereens, 2000; LeFevre, 2013; Marshall, 1998; Stachler, Young, & Borr, 2013).

Evaluation theoretically causes teachers to change their classroom instruction and drives professional development for the district. According to Bereens (2000), sustainable and effective change in classroom practice is predicated upon the ability of teachers to trust their leaders. Trust, as it relates to the educational process of evaluation, is frustrating for both leaders and teachers (Bereens, 2000).

Trust is also an ingredient when change occurs. LeFevre (2013) utilized an analytical framework of risk to explain why educational change generally fails. Risk-taking and educational change were determined to be symbiotic. Acknowledging crucial stages and elements related to educational change, the presence of trust in the organization was determined to be critical. According to Marshall (1998), teachers need to trust feedback from their evaluation and not view them as criticisms. Trust carries over into professional development, as teacher evaluation drives instructional change and it also serves to guide professional development for the district.
School leaders are best poised to assess teacher effectiveness and quality through performance evaluations (Kimball & Milanowski, 2009). Kimball and Milanowski (2009) conducted a study with the hope that the results could lead to improved teacher evaluation practices. Their study was designed to better understand evaluator decision making and to learn whether differences in decision-making could help account for the differential validity they had observed in principal evaluations of classroom performance.

Kimball and Milanowski (2009) conducted a sequential mixed-methods research study that focused on 23 school leaders in a large public school district in the United States with 3,300 teachers. This district had employed Danielson’s (2010b) standards-based evaluation system for 3 years prior to the study, and it identified differences in principals’ (a) motivation to be accurate, (b) knowledge of instruction and evaluation, and (c) school context. The study sought to uncover if motivation for accuracy, knowledge and skill, and school context could explain why some evaluators’ ratings were correlated to student achievement and others were not.

The results identified substantial variation across evaluators in terms of strength and direction of the relationship between teacher evaluation scores and student achievement scores (Kimball & Milanowski, 2009). The 2001 to 2002 school year identified 28% of the evaluators scoring low in the validity category. In comparison, 44% of the evaluators were identified for the same school year as falling into the high validity category. This trend was found for all school years involved in the study.

According to the study, motivation, knowledge of the course content, understanding of evaluation protocols, and desire to maintain collegial relationships with teachers resulted in discrepancies between evaluators (Kimball & Milanowski, 2009).
They conclude the study by cautioning that reliability of scores among evaluators will vary unless extensive training on standards-based criterion evaluation systems occurs. Robinson (2011) affirmed this caution by noting that the lens through which an evaluator views teaching quality is one that encompasses personal values. Subjectivity on the part of the evaluator influences observation practices. This factor is one that researchers note as evident in practice, with extensive training in observation protocols as the key to controlling it (Kimball & Milanowski, 2009; Welsh, 2011). Namaghi (2010) acknowledged, “evaluation can be limiting if it is judgmental” (para. 6). Focusing on positive elements of teaching quality increases the likelihood of internalization and reform of practice (Namaghi, 2010).

To assist in controlling for subjectivity, implementation of standards-based assessment rubrics is advocated (Danielson, 2010b). Standards offer a common language that describes instructional practices so that the teacher and evaluator are able to conduct deep and meaningful conversations about student learning (Danielson, 2010a). An overall goal of teacher observation is to improve student achievement (Holtzapple, 2003). Holtzapple (2003) contended that standards-based teacher evaluations, therefore, have the potential to improve student learning. The ensuing study assessed the relationship among teachers’ evaluation scores and student achievement gains, as measured by a proficiency-based standardized test.

Holtzapple (2003) obtained data for the study from Kimball and Milanowski’s 2009 study that focused on the Cincinnati Public School Teacher Evaluation System. Charlotte Danielson’s (2011) *The Framework for Teachers* was of particular interest for the study. Data included teachers’ scores on the individual rubric domains or the final
composite score that was the sum of all four domain scores. Domain three focused on classroom instructional strategies and was of particular interest because it had the greatest potential to affect student achievement. Initial research was concerned with the relationship between domain three, classroom instruction, and student achievement (Holtzapple, 2003). It was surmised that if teachers scored highly on this domain, it would transfer to higher student achievement on the standardized tests.

Results indicated that the evaluation scores were significantly lower the second year of implementation compared to the first year (Holtzapple, 2003). One possible cause for the variance in evaluation scores included a 20% increase in first-year teachers in the 2001-2002 school year. Teachers’ scores were categorized as unsatisfactory, basic, proficient, and distinguished. Students whose teachers scored unsatisfactory or basic had lower gains for both years compared to their peers whose teachers scored in the higher categories. In years 2001 to 2002, the levels were 10.88 at unsatisfactory, 14.36 for basic, 14.60 for proficient, and 14.60 for distinguished. The second year found lower evaluation scores for the teachers as well as lower residuals when combined with student achievement.

Holtzapple (2003) concluded that teachers who scored in the lower tiers of the evaluation system, such as unsatisfactory and basic, negatively impacted student achievement because the students scored lower than their peers on state and district tests. This was conversely true for teachers who scored highly on their evaluations. Their students scored well on district and state tests. In both cases, teachers’ scores were related to student scores. Confounding factors may have played a role in the scoring of teacher evaluations, therefore, Holtzapple (2003) noted that the study was only sensitive
to identifying the two extremes of the evaluation process: poor and good. The study emphasized the impact of teacher quality as a key factor in predicting student achievement.

Kimball and Milanowski (2009) offered will, skill, and motivation as factors that influence teacher evaluation, and Darling-Hammond (2000) furthered that variables assessing teacher quality are more indicative of student success than are student demographics. Mashburn, Meyer, Allen, and Pianta (2013) furthered this inquiry of teacher evaluation by conducting an experimental study on the effect of different observational procedures on score reliability and validity. Their goal was to identify observation procedures that maximized score reliability and the validity of score inferences.

Mashburn et al. (2013) began the study by acknowledging that one or more trained-rater observed a classroom for a specified period and then provided feedback, as a basic characteristic of most observation systems. Another key component of most evaluation systems is the need to provide additional information to evaluators beyond what is observable in the classroom. For this study, the researchers focused on classroom observational aspects of teacher evaluation.

The methodology for the study included review of data obtained from teachers and students in Grades 6-11 who were participating in another study regarding secondary school teachers (Mashburn et al., 2013). The study involved 100 teachers from eight schools located in the southeastern part of the United States. Random assignment placed 47 teachers in the treatment condition and 43 teachers in the controlled condition. Teachers videotaped 40-minute lessons and sent them to the researchers within pre-
determined windows of time during the year. Only teachers who submitted all required tapes, with each tape being audible and without technical issues, were utilized. The final tally included 141-videotaped lessons from 47 teachers. Teacher demographic information and student achievement scores were also included as data for the study.

Mashburn et al. (2013) designed the experimental conditions by varying the observation length of the lessons. Lessons were rated using the Classroom Scoring Assessment System Secondary (CLASS-S) instrument to assess teacher quality for this study (Pianta et al., 2008). The conditions for the study included:

1. Raters judged a single 40-minute lesson; 1 x 40 model.
2. The lesson was divided in half, and two raters observed the first 20 minutes, rated the quality level, and then rated another random 20-minute lesson segment; 2 x 20 random model.
3. The lesson was divided in half and two raters observed the first 20 minutes, rated the quality level, and then rated the second half of the same lesson; 2 x 20 ordered model.
4. The lesson was divided into four 10-minute segments, and four raters observed 10-minute segments, and then they rated the quality before moving on to the next segment. They viewed the entire lesson sequentially in 10-minute blocks of time; 1 x 40 model.

Mashburn et al. (2013) employed three equations for computing the number of lessons and segments viewed along with the time allotments. They determined that increasing the number of lessons always resulted in lower relative error variance than increasing the number of segments (Mashburn et al., 2013). Increasing the number of
raters was not as effective in lowering relative error variance as increasing the number of lessons. Of significance was the overall finding that lower relative error variances occurred when rating multiple segments of a lesson rather than using one 40-minute lesson.

Results from the study (Mashburn et al., 2013) indicate that while the shortest and most frequent manner of classroom observation (4 x 10 model) produced the lowest generalizability coefficients, it also posted the lowest universe score variances. Short 10-minute observations may not provide adequate time for observers to evaluate the lesson with fidelity. Single, 40-minute lesson observations found large relative error variance and low validity coefficients. High correlation and high generalizability estimates are attributed to the 20-minute segmented lessons. Ordered or random viewing of lessons did not appreciably change the results in regard to generalizability.

Mashburn et al. (2013) concluded the study, acknowledging operational procedures related to the length of observation can affect the reliability and validity of the scores. Teachers’ evaluation scores are used by some districts for pay and promotion decisions as well as assignment of courses. Ensuring reliable and valid evaluation procedures and operations are critical for fair and reliable evaluation reports. Researchers have acknowledged an inherent predisposition to bias when conducting performance appraisals, thus negating reliable and fair evaluation procedures.

**Evaluator bias.** Research acknowledges instances where bias has the potential of occurring during classroom observations and, by default, questions the validity of assigned scores (Bell et al., 2012; Hill & Grossman, 2013). Time of day, content-knowledge, experience evaluating, mood, and physical health of the evaluator may
unintentionally result in a biased score. Scoring methodology does not account for these potential instances of bias (Bell et al., 2012).

Inter-rater reliability is an area where bias has been identified (Hill & Grossman, 2013; Rockoff & Speroni, 2010). Rockoff and Speroni (2010) noted that “application of standards can vary significantly across individuals responsible for making evaluations and implementation of any evaluation system should address this issue” (p. 264). Inconsistency between evaluators is a concern for both administration and teachers. To address inter-rater reliability, New York State requires yearly training for evaluators (EngageNY, 2013).

Bell et al. (2012) maintains that consistency between and within evaluators is imperative and that scoring should not fluctuate between observers. They further contend the threat of bias when only one observer scores a given teacher. Content knowledge, situational issues between teacher and evaluators, as well as pre-conceived perceptions regarding content also have potential for bias (Hill & Grossman, 2013). Bias variables outside of a teacher’s control, which result in an observation score affecting tenure or merit pay, are cause for concern (Bell et al., 2012). Evaluating teachers and assigning scores is not an exact science. Research has documented that subjectivity on the part of the evaluator as a concern affecting overall teacher evaluations (Hill & Grossman, 2013).

Similarly, Borman and Kimball (2005) conducted a quantitative study in the second largest school district in Nevada, Washoe County School District. Using standards-based evaluation ratings for close to 400 teachers and results for over 7,000 students Grades 4-6, they investigated the distribution of achievement effects of teach
quality. Washoe County was experiencing rapid growth, which resulted in 400 teachers hired annually.

Student demographic data provided construct variables for minority and free and reduced-fee lunches (Borman & Kimball, 2005). To measure proficiency, they utilized state and district assessments that measured the Nevada State content and performance standards for Grades 3-6. Composite scores of teachers’ performances associated with the art of teaching and delivery were incorporated into the study.

Borman and Kimball (2005) identified Grade 4 reading teachers with higher evaluation scores made some progress in closing the achievement gap that separated poverty and non-poverty students. To a lesser extent, it also occurred with Grade 4 math teachers with high evaluation scores. They posited that bias was present in the assignment of courses and students to teachers. According to their study, evaluators exhibited a positive bias toward teachers with higher achieving students, than those with challenging classes. Classes with higher concentrations of poverty and minority students were identified as challenging classes. The relationship between teacher quality and student assignment found higher performing students assigned to highly effective teachers (Adamson & Darling-Hammond, 2012; Borman & Kimball, 2005).

According to Borman and Kimball (2005), scoring of teacher quality appears related to what courses and students are assigned to the teacher. They further contended that perception of teaching quality may find bias in how teacher performance is scored. Bias was noted as grounds for further inquiry. Another area noted as a concern was the definition of teacher quality. Closing the achievement gap is an important aspect of the
evaluation system, and thus, a clear definition of high-quality teachers needs to be more refined.

A preconceived notion that teachers of lower achieving students are, themselves, less effective (Borman & Kimball, 2005) may be viewed as a form of evaluator bias. Beare (2014) conducted a qualitative study that examined the effect of teacher socioeconomic status and ethnicity on principals’ evaluations of teachers’ preparation. Questionnaires distributed to supervising principals by the California State University System from 2006 to 2010 provided data for this study. The college application process provided demographic data for the study. A series of analyses of variance were performed with candidates grouped based upon their demographic variables. Demographic variables included family income was determined by combining parent incomes, parental education level, and ethnicity.

Income levels of parents, when compared to the success of their children did not reveal any statistical significance (Beare, 2014). Parental education was divided into seven categories ranging from no high school through postgraduate degree. No statistical significance was determined for parental income and the rating for graduates from teacher-preparation programs.

Beare (2014) acknowledged that principals do not show bias in rating the preparation of teacher-education graduates based on demographic factors. Family demographics, such as parental educational levels, income, and ethnicity, did not have an effect on how evaluators rated recent graduates from teacher-education programs. It was determined that principals rated based on what they observed in the classroom and not on extraneous factors. This study acknowledged a lack of socioeconomic and ethnicity bias.
Bell et al. (2012) concurred with the premise of studying contextual features in an effort to remove bias, especially if certification decisions and merit pay are based upon the evaluation results.

Bell et al. (2012) contended that teacher evaluation informs professional development and making personnel decisions. Borman and Kimball’s (2005) study supports this premise as their evidence supported the hypothesis that low-performing teachers were assigned low-achieving students, and conversely the same was true for high-performing teachers and students. Bell et al. (2012) developed a validity argument approach for use on observation protocols used to assess teacher quality. Data for the study were drawn from a validity study on the CLASS for secondary classrooms. Data from 82 algebra classrooms was the foundation upon which they regressed contextual factors to build a validity argument for observation protocols. Observation scores, value-added models, generalizability studies, and measures of teacher knowledge, student achievement, and teacher and student beliefs comprised the contextual factors for the study.

A clarification between teaching and teacher quality preceded the validity study. Teaching quality was defined as the interaction between teachers and students within the classroom, whereas teacher quality was defined as the quality of these interactions that are teacher driven (Bell et al., 2012). Teaching quality comprises many facets and it is not possible to isolate them. Figure 2.1 identifies a conceptualization model as put forth by Bell et al. (2012) concerning teacher quality.
Figure 2.1. The contextual factors, constructs, and measures associated with teaching quality. Adapted from “An Argument Approach to Observation Protocol Validity,” by C. A. Bell, D. H. Gitomer, D. F. McCaffrey, B. K. Hamre, R. C. Pianta, & Y. Qi, 2012. Educational Assessment, 17(2), 62-87, p. 64. Copyright 2012 by the Taylor & Francis Group, LLC.

Development of the validity argument included describing observation systems as a set of dimensions of teaching (Bell et al., 2012). Charlotte Danielson’s (2011) The Framework for Teaching rubric and the CLASS are two such examples of observation rubrics. Protocols for observations are negotiated within school districts, and they include the number of observations that will occur during the year, the standards-based rubric that will be employed, and a description of how the scoring will occur.

Observation scores connect the observed lessons with the intentions on the rubric, which brings into question the accuracy of scores. Bell et al. (2012) addressed accuracy with the introduction of inter-rater reliability: a lesson is scored the same—no matter who
observed and scored the lesson. Inter-rater reliability is prone to be affected by bias. Bias, according to Bell et al. (2012), has two principle sources, the manner in which observers are assigned to lessons, and how the observers interpret the scoring tool. Bias may take the positive form where there is greater leniency, or it may be negative where scores are conservatively assigned. Bias is an issue that the researchers clearly outlined as an area of concern that required further study.

Data for the application of the validity approach, the Toward an Understanding of Classroom Context (TUCC), was drawn from a single study in 2009 to 2010. The study took place in a large, urban, mid-Atlantic school district. The TUCC study employed the CLASS research-based observation protocol that has three domains: emotional support, classroom organization, and student engagement (Bell et al., 2012). Participation included 82 algebra teachers, all who had at least a bachelor’s degree, and 53% had a master’s degree. The teachers were observed by multiple observers while they were teaching four or five full-length lessons. Observers, in turn, completed a weekly calibration exercise to ensure inter-rater reliability. Students were offered a pre- and post-test and the scores were entered for student achievement data.

Descriptive statistics were employed to present data from the qualitative study. Bell et al. (2012) recorded that observers’ scores were within one point of the master observers on 70% of the lessons for all dimensions, 80% of the lessons for four dimensions, and 90% of the lessons for two dimensions. On double-scored lessons, observers’ agreement was between 65 and 97%. A study limitation was that bias was not able to be determined based on observer assignment. However, assignment of scores
found bias in that there were significant differences in how stringent or lenient observers were.

The validity argument approach makes it evident that explicit understandings of the relationships among constructs, instruments, and inferences needs to occur before evaluation results are used for human resource decisions (Bell et al., 2012). Deciding who will observe what lessons, when training will occur, how often, and what protocols will be in place to ensure inter-rater reliability are issues that require immediate attention. These issues are relevant and timely, given federal and state initiatives to hold teachers more accountable for student achievement. Teachers, in the meantime, are experiencing both positive and negative responses to the feedback generated from the move to standards-based observations. How teachers perceive the feedback, whether positively or negatively, determines their perception of the overall message (Pearce & Porter, 1986).

Pearce and Porter (1986) hypothesized that “attitudes toward the performance appraisal systems and organizational commitment will decrease and remain lower for those receiving satisfactory ratings, whereas the attitudes of those receiving higher appraisal ratings will remain unchanged” (p. 211). Their quasi-experimental study investigated the attitudinal impacts of the receipt of formal performance appraisal feedback. As part of the research on the effects of the Civil Service Reform Act, the study utilized data that was collected before and after implementation of the new appraisal system. Employees from the National Aeronautics and Space Administration research center and the Department of Defense were included in the study. This was the first time the agencies were assigning a performance rating to the evaluations, from which promotions, pay increases, and layoffs would be based (Pearce & Porter, 1986).
Employees were rated on objectives, and these ratings were summarized through an approved formula to determine final composite ratings (Pearce & Porter, 1986). The new evaluation process mandated face-to-face feedback sessions where the subordinate signed the appraisal. The study employed a time-series, repeated-measures, quasi-experimental design, and it included an attitude questionnaire covering a 30-month timeframe. The questionnaires were distributed three times prior to dissemination of the ratings, then two months after the ratings were released, and again two months later. The respondents were randomly selected for the anonymous survey that received an 81% response rate.

Satisfactory ratings between time one and two for managers elicited a significant drop in attitudes for both organizations, with no change in attitudes for highly rated managers (Pearce & Porter, 1986). Both groups also viewed the system more positively between Time 2 and Time 3. The research center found attitudes dropped for satisfactorily rated managers between Time 1 and Time 2, and both groups were significantly more positive after the second feedback sessions. The findings for the aeronautic organization supported the hypothesis as attitudes dropped with a satisfactory rating between Time 1 and Time 2, and remained so a year later. In contrast, the aeronautical engineers held positive attitudes throughout the entire process.

Attitudinal changes were more apparent at the research center (Pearce & Porter, 1986). Aeronautical engineers did not display differences over time nor between ratings of satisfactory or high. The results supported the second hypothesis: low ratings resulted in a significant drop in attitudes toward the organization. Unintended consequences from the appraisal process, when raters were trying to be precise, had the potential to affect
overall organizational commitment. Systems that rank employees need to be mindful that this process is likely to generate the loss of positive feelings. Organizational citizenship behavior reflects feelings of employees, both positively and negatively.

Porter and Porter (1986) stated that attitudes of all employees, those at the top, and those identified as solid employees, need to be considered when designing and implementing performance appraisal systems.

Overall evaluation of teachers sometimes includes the use of student performance data. Milanowski (2004) studied the relationship between teacher evaluation scores and student achievement on district and state tests in Grades 3-8 math, reading, and science from the 2001 to 2002 school year. Correlations varied between grade levels considerably when compared with teacher evaluation scores.

Based upon the results, there was a small degree of validity between teacher evaluation scores and student assessments. The teachers were identified as who had a greater impact on student achievement. Identifying teachers whose students are more successful on assessments will enable districts to distinguish teaching practices that are more effective for learning.

**Administrator impact on the evaluation process.** Administrators affect the overall process of the evaluation system in regard to the willingness of teachers to enact suggested changes resulting from feedback from the overall appraisal system (Danielson, 2010b; Pearce & Porter, 1986; Runhaar, Sanders, & Yang, 2010; Whiting et al., 2007). Runhaar et al. (2010) investigated how teachers’ reflections and request for feedback are related to occupational self-efficacy, learning goal orientation, and transformational leadership.
Runhaar et al. (2010) affirmed the connection between a teacher’s motivation to change and his or her perceived opportunity to do so. Perception is developed through a fairness process and trust in leadership. The interplay between leadership and teachers in regard to evaluation is critical at the feedback stage. Leadership that encourages reflection and acknowledges the relationship between reflection and practice enhances self-efficacy (Runhaar et al., 2010). Wahlstrom and Louis (2008) identified self-efficacy and trust in leadership as strong predictors of instructional behavior in the classroom.

Wahlstrom and Louis (2008) conducted a quantitative study that examined factors present in teacher-principal interactions that related to instructional practices by the classroom teacher. Surveys were completed by a sample of 4,165 teachers from across the US. Dependent variables included standard contemporary practice, focused instruction, and flexible grouping practices. The independent variables included principal trust and shared leadership. Linear regression was employed to address the research questions.

The results identified that trust in school principal leadership was less significant than shared leadership (Wahlstrom, 2008). Principal’s sharing leadership decisions with non-administrative faculty members was determined to lead to a greater degree of focused instruction than trust in the principal. These results were in contrast to later studies by Namaghi (2010) and LeFevre (2013) who stated that trust is critical in maximizing change in a classroom teacher’s instructional practice.

Wahlstrom (2008) determined that the presence of shared leadership and professional learning communities impact a teacher’s implementation of research-based instructional practices, focused instruction, and flexible grouping practices. Trust was
less of a factor than initially expected (Wahlstrom, 2008). Principals are instructional leaders within their schools, and the scope of their approach to improving instruction varies among districts and schools (May & Supovitz, 2011).

May and Supovitz (2011) focused their study on the scope of principals’ efforts to improve instruction. Part of a larger longitudinal study, from 2005 through 2007, the qualitative study utilized nominal self-reported data in the form of web-based logs and surveys. Principals completed daily web-based logs that identified time spent across nine categories: building operations, finances, community and parents, school district functions, students, personnel issues, goals, instructional leadership, and professional growth. Of the teachers who identified themselves as having professional and ongoing dialogues with their principals, 1,608 completed an annual five-point Likert scale survey. Included on the survey were five items:

1. the principal and the teacher discussed the teacher’s instruction;
2. the principal observed the teacher instructing a class;
3. the teacher observed the principal instructing a class;
4. the principal provided feedback after observing the teacher’s instruction; and
5. the principal reviewed the work produced by a teacher’s students (May & Supovitz, 2011, p. 340).

Results identified that school principals spent an average of 2 to 5 hours per week on activities associated with instructional leadership (May & Supovitz, 2011). Broad and targeted instructional leadership by principals with teachers varied widely without regard to school size, and broad and targeted instructional leadership did not appear to have a significant effect on the change in instruction. However, a small increase was found in
reading/ELA teachers in the same school. May and Supovitz (2011) posited that this increase may be the result of principals targeting a small subset of teachers. Teachers who reported having regular interaction with their principals also reported higher changes in instructional practices.

May and Supovitz (2011) concluded that principals were more effective at eliciting instructional changes with teachers when they were able to target focused time with them. This ability to change practice through targeted leadership did not vary by district size. What was significant was the time spent by principals with individual teachers. A principal’s choice of how to allocate his or her time was noted as being critical for his or her role as an instructional leader.

**Evaluation systems.** School districts in New York State have options regarding how they assess instructional performance of teachers. Most districts employ direct classroom observations of varying length as well as the use of teacher-generated portfolios. Attinello, Lare, and Waters (2006) examined the value of district-wide portfolio-based teacher evaluation systems, and they found support for this authentic facet of evaluation.

A mixed-methods study was conducted in the southeast region of the United States in a public school district comprising 20,000 students enrolled, 752 teachers, and 23 school buildings. A portfolio-based system for evaluation had been in place for 4 years. Attinello et al. (2006) conducted a qualitative study by administering surveys developed by Tucker, Stronge, and Gareis in 2002 entitled Teacher Survey: Perceived Value of Teacher Portfolios along with the Administrator Survey: Perceived Value of
Teacher Portfolios. Every administrator and teacher received the survey, and the return response rate was 63.4% teachers and 70.8% for administrators.

Descriptive statistics revealed both the administrators and teachers agreed that portfolios are an authentic form of assessment (Attinello et al, 2006). They did not find significant differences between perceived accuracy, a comprehensive picture of teaching performance, nor impact on professional development between administrators and teachers. It was determined from the study that portfolios allow teachers to have a more active role in their performance evaluation. Administrators acknowledged the increase in collaboration and sharing among staff through the use of portfolios.

Attinello et al. (2006) concluded that the use of portfolios for performance appraisals shows promise for ensuring accuracy through a comprehensive format. Teachers enjoyed the use of this system, although they acknowledged the need for multiple methods for a more accurate measure of performance. Portfolios would be a difficult tool for teachers to use when illustrating use of technology in the classroom. While technology is included as a component in the common-core teaching standards, it is not always an element within teacher evaluation tools.

Whale (2006) studied the degree to which public schools in Michigan included teacher technology skills in their evaluations. The qualitative study employed descriptive statistics to analyze data from self-reported and coded surveys. Surveys were sent to every elementary, middle school, and high school teacher in Michigan. Final participation included 220 teachers from 39% of the state’s school districts. Nominal variables included internet, technology, software, hardware, multi-media, and computers. The findings revealed 19.1% of districts included teacher technology skills in
observations (Whale, 2006). Further results identified 81% of districts omitting teacher technology skills from evaluations. Direct reference of technology standards was not specifically identified, but the technology standards were included 72.5% of the time within the instruction section of teacher evaluation rubrics.

There is conclusive evidence that technology affects student learning, and is an important pedagogical tool for educators (Whale, 2006). Results from the study indicate that technology is something teachers engage in, but they do not receive much feedback from evaluators, as this is not directly included in the teacher evaluation processes.

**New York State annual professional performance review program.** The U.S. Education Department awarded New York State $696,646,000 from the federal Race to the Top Competition, August 2010 (NYSED, 2010). In compliance, New York State passed Education Law §3012-c, which changed the conduct of the APPR for teachers (EngageNY, 2013). Annual evaluations of all teachers are now legally mandated for all public and charter schools in New York State (EngageNY, 2013). Inclusion of student achievement and growth scores are a new component added to the teacher evaluation process. Accountability for student academic performance is the guiding principle behind the addition of student achievement to teacher the APPR process.

Mandated annual performance evaluations for all teachers is the foundation for the overall APPR process (EngageNY, 2013b). APPR utilizes a 100-point scoring scale. From 2012-2013 through the 2014-2015 school year, multiple classroom evaluations accounted for 60 points, student growth scores 20 points, and local measures that assess either growth or achievement of student learning determined the remaining 20 points. Math and English teachers in Grades 3-8 are provided a 0- to 20-point growth score for
their students. Computation of the score occurred by assessing the growth of individual students between annual tests in English and in math. All other teachers earned 0 to 20 points by developing district-approved student learning objectives (SLOs) (OCM BOCES, 2011). Locally selected measures of student achievement comprised 0 to 20 points for all teachers. School districts chose from state-approved, third-party assessments for the local measures, or they developed their own. Therefore, 20 points were accrued through state provided growth scores or SLO’s and 20 points were accrued through locally selected measures of student achievement or growth. A total of 40 points are awarded based upon student learning.

APPR met with significant change as the Governor of New York signed into law a revised evaluation system for teachers and principals on April 13, 2015 (NYSED, 2015b). The new law, §3012-d, requires teachers be evaluated on two categories, student performance and teacher evaluations. With this change, both student achievement and teacher evaluation comprise 50% of the composite score (NYSED, 2015b). The change effectively removed portfolios of teacher artifacts and goal-setting, and it mandates multiple raters for each teacher. Effectively, the principal must evaluate each teacher, and the resulting score must comprise at least 80% of the overall evaluation score. In addition, a certified lead evaluator who is not assigned to the school building, known as an independent rater, must evaluate each teacher. The independent evaluator’s score must not exceed 20% of the teacher’s evaluation score.

Districts must work with teacher unions to collectively bargain for the choice of a state-approved rubric for the evaluation points (EngageNY, 2013b). Multiple classroom
observations comprise 50% of the final composite score (NYSED, 2015a). Negotiated standards-based rubrics are the basis for scoring classroom observations.

Teaching standards clearly identify what teachers should know and be able to do (Danielson, 2010b). Standards-based systems measure teachers’ practice against the concrete standard. Danielson (2010b) identified four main standards: (a) planning and preparation, (b) the classroom environment, (c) instruction, and (d) professional responsibilities. The New York State United Teachers 2012 standards rubric utilizes: (a) knowledge of students and student learning, (b) knowledge of content and instructional planning, (c) instructional practice, (d) learning environment, (e) assessment for student learning, (f) professional responsibilities and collaboration, and (g) professional growth (EngageNY, 2013b). Standards-based rubrics are scored on a continuum that spans from highly proficient through not observed. Frequent application of standards-based observation, coupled with specific and timely feedback to the teacher, is significantly related to gains in student achievement (Darling-Hammond, 2013).

Classroom instruction and student achievement result in an annual quality rating and a final composite score for all teachers. Highly effective, effective, developing, and ineffective comprise the four quality rating categories. Composite scores and annual quality ratings are new elements for public school teachers in New York State.

By September 1, 2011, public school districts across New York State adopted APPR plans for their instructional staff (EngageNY, 2012b). Initially, the law only covered common-branch teachers: English Language Arts (ELA) and math teachers in Grades 3-8. Effective in the 2012-2013 school year, all instructional teachers were included in the revised APPR process. APPR law §3012-d requires districts to negotiate
and implement APPR plans for the 2015-2016 school year (NYSED, 2015a). School districts unable to successfully implement negotiated plans may apply for a hardship waiver in November if they can present good-faith efforts on the part of the district and the teacher’s union. Should a plan not be in place by March 2016, a second and final hardship waiver may be applied for that will close out the 2015-2016 school year.

Mandated by law, school districts must be fully operational under §3012-d by the start of the 2016-2017 school year.

Table 2.1

Permissible Statewide Ranges For Teacher Evaluations

<table>
<thead>
<tr>
<th>Category</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Effective</td>
<td>3.50-3.75</td>
<td>4.00</td>
</tr>
<tr>
<td>Effective</td>
<td>2.50-2.75</td>
<td>3.49-3.74</td>
</tr>
<tr>
<td>Developing</td>
<td>1.50-1.75</td>
<td>2.49-2.74</td>
</tr>
<tr>
<td>Ineffective</td>
<td>0</td>
<td>1.49-1.74</td>
</tr>
</tbody>
</table>

*Note.* The final composite scores for teachers employs the use of a matrix that takes into account the category rating and not the score (NYSED, 2015a).

A matrix was utilized to determine a teacher’s quality rating. As part of the matrix, evaluators utilized rubrics to assign scores of 0-4 for the prescribed indicators (NYSED, 2015b). Actual cut scores for determining highly effective, effective, developing, or ineffective occurred locally using the parameters in Table 2.1.
Table 2.2

*State Prescribed Conversion Table Assigning Student Growth Scores*

<table>
<thead>
<tr>
<th>Rating</th>
<th>SLOs</th>
<th>Scoring Range</th>
<th>State-Provided Growth Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent of students meeting target (%)</td>
<td>MGP Range</td>
<td>Rating</td>
</tr>
<tr>
<td>I</td>
<td>0-4</td>
<td>0</td>
<td>3-23</td>
</tr>
<tr>
<td>I</td>
<td>5-8</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>I</td>
<td>9-12</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>I</td>
<td>13-16</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>I</td>
<td>17-20</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>I</td>
<td>21-24</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>I</td>
<td>25-28</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>I</td>
<td>29-33</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>I</td>
<td>34-38</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>I</td>
<td>39-43</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td>I</td>
<td>44-48</td>
<td>10</td>
<td>33</td>
</tr>
<tr>
<td>I</td>
<td>49-54</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>I</td>
<td>55-59</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>D</td>
<td>60-66</td>
<td>13</td>
<td>29-37</td>
</tr>
<tr>
<td>D</td>
<td>67-74</td>
<td>14</td>
<td>38-4*</td>
</tr>
<tr>
<td>E</td>
<td>75-79</td>
<td>15</td>
<td>36-48*</td>
</tr>
<tr>
<td>E</td>
<td>80-84</td>
<td>16</td>
<td>49-55</td>
</tr>
<tr>
<td>E</td>
<td>85-89</td>
<td>17</td>
<td>56-68</td>
</tr>
<tr>
<td>H</td>
<td>90-92</td>
<td>18</td>
<td>67-68*</td>
</tr>
<tr>
<td>H</td>
<td>93-96</td>
<td>19</td>
<td>69-72</td>
</tr>
<tr>
<td>H</td>
<td>97-100</td>
<td>2</td>
<td>73-94</td>
</tr>
</tbody>
</table>


The New York State Education Department issued a chart that breaks down student growth scores for teachers, or the equivalent student learning objectives, into a similar categorical breakdown for teachers (Table 2.2). Table 2.3 illustrates the matrix for determining a teacher’s final APPR rating.
Table 2.3

*The Matrix*

<table>
<thead>
<tr>
<th>Student Performance</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Effective (H)</td>
<td>H</td>
</tr>
<tr>
<td>Effective (E)</td>
<td>H</td>
</tr>
<tr>
<td>Developing (D)</td>
<td>E</td>
</tr>
<tr>
<td>Ineffective (I)</td>
<td>D</td>
</tr>
<tr>
<td>Highly Effective (H)</td>
<td>H</td>
</tr>
<tr>
<td>Effective (E)</td>
<td>E</td>
</tr>
<tr>
<td>Developing (D)</td>
<td>E</td>
</tr>
<tr>
<td>Ineffective (I)</td>
<td>D</td>
</tr>
<tr>
<td>Effective (E)</td>
<td>E</td>
</tr>
<tr>
<td>Developing (D)</td>
<td>D</td>
</tr>
<tr>
<td>Ineffective (I)</td>
<td>I</td>
</tr>
<tr>
<td>Ineffective (I)</td>
<td>I</td>
</tr>
</tbody>
</table>

*Note.* Growth or other comparable measures are inclusive of the state-provided score or the use of student learning objectives (NYSED, 2015a).

John King, the New York State Commissioner of Education through the 2015 school year, noted that implementation of the revised APPR serves three purposes: (a) it focuses on student learning, (b) it helps teachers understand their impact on students, and (c) it provides teachers with meaningful opportunities for professional growth (EngageNY, 2013). The Common Core Learning Standard (CCLS) focuses on student learning. Embedded in the standards-based rubrics are the CCLS that districts may choose from as a basis for their evaluation systems. These learning standards also serve as the basis for mandated state assessments in Grades 3-8, and soon they will include all high school Regents exams.

Student assessment data in Grades 3-8 will result in a state-provided growth score that comprises 50% of a teacher’s composite APPR score (NYSED, 2015a). Under the initial implementation of APPR, the state-provided growth score comprises 20% of a
teacher’s composite APPR score. Linking teacher evaluation to student performance has caused controversy. NYSUT is a 600,000-member union organization for teachers in New York State (NYSUT, 2014). In June 2011, NYSUT sued the New York State Board of Regents over APPR regulations. Casagrande (2011) posited that:

It takes something extremely important for NYSUT to sue the Board of Regents. It happened only once in NYSUT’s 40 years’ existence. But in June, NYSUT was forced to sue the Regents again, concerning the regulations adopted by the Regents to implement Annual Professional Performance Review under NYS Education Law §3012-c. On Aug. 24, the Albany County Supreme Court ruled largely for NYSUT in a lawsuit that should have been unnecessary, had the regulations followed the letter and intent of the law. (Casagrande, 2011, para. 1)

Incorporation of student growth scores based upon standardized testing was the basis for most of the suit (Calabrese, 2011). The law specifically states that student growth or achievement scores should comprise 20% of a teacher’s final composite score. The New York State Education Department and the New York State Board of Regents changed the percentage to 25%. The judge also upheld the complaint that implementation of accountability procedures violated collective bargaining mandates. NYSUT adamantly stated that it supports the APPR statute, and that this suit was filed to ensure consistent application of the tenets of the law with resulting mandates (Calabrese, 2011).

Issues of teacher privacy arose as composite scores were calculated and then submitted to the New York State Education Department. On February 28, 2012, the New York Post identified a New York City teacher as being the worst teacher in the city (Pathe
The “name, salary, and a link to a database of value-added scores for teachers across the city” was also included in the *New York Post* story, according to Pathe and Choe (2013). Posting of this information caused concern over the privacy rights of public school teachers. The GAO (2013) included this concern in their report:

> District and union officials in New York said the release of teacher evaluation ratings to parents added to concerns about evaluation systems. Officials in one small district said their teachers were particularly concerned because protecting their anonymity might be difficult even if data are aggregated and not linked to individual teachers. (GAO, 2013, p. 19)

Who will be able to access a teacher’s composite score and the HEDI label is of concern for educators. In March 2014, state education officials admitted problems with the new teacher ratings and the APPR process (Hildebrand, 2014). The March 2014 article posited that NYSED was still working on the final report of teacher evaluations from the 2012-2013 school year. Data from school districts regarding teacher evaluation were due to the NYSED in October 2013 (EngageNY, 2013). Hildebrand (2014) reported that state officials acknowledged flaws existed regarding the APPR composite scores and the HEDI labeling of teachers. Hildebrand (2014) concluded with a quote from state education official, David Albert: “Some growing pains were to be expected in an initiative as complex and extensive as the state’s evaluation system” (Hildebrand, 2014, para. 23).

**APPR implementation.** Public school districts throughout New York State were remanded with the task of having approved APPR plans that were in place for the 2012-2013 school year (EngageNY, 2012a). Developing APPR plans proved challenging as
exemplars were not available for reference. School districts established committees of teachers and district administrators to negotiate the terms of the APPR agreements. Teachers missed time in the classroom and incurred additional pay from districts for APPR negotiation meetings outside of the school day. Time out of the classroom required districts to pay for substitute teachers. Thus, the cost of implementing APPR plans included tangible and intangible elements. Payment for substitutes was a tangible cost, and intangible costs included teachers missing class time for meetings and administrators missing time from other administrative duties (W. Furlong, personal communication, April 28, 2015). These costs were not subsidized by NYSED and were incurred by the districts themselves.

With the newly revised §3012-d APPR law, the initial challenges to implementation remain. NYSED acknowledges the challenges districts face and has allowed for extensions to be filed in order to have solid collectively bargained APPR plans in place (NYSED, 2015b). Extensions, known as hardship waivers, are available in November 2015 and March 2016 for the 2015-2016 school year when districts are able to prove good faith in negotiating a plan. Hardship waivers are not permissible beyond the 2015-2016 school year (NYSED, 2015b).

Time spent preparing evidence for evaluations tax classroom teachers. Everard (W. Furlong, personal communication, April 27, 2015) acknowledged spending 24 hours preparing her APPR portfolio—hours that could have been spent preparing for her students. Mirroring Everard’s sentiment, Schug (W. Furlong, personal communication, April 27, 2015) stated spending a considerable amount of time on APPR in lieu of working on classroom material. She furthered that time spent on APPR was time that
would be better served helping students. These intangible costs are more confounding when calculating the equation for final APPR scores. Teacher evaluations comprise up to 60% of the scores with 20% from local measures and the last 20% determined by student test scores on state exams.

Another tangible cost associated with the APPR process includes maintenance of the myriad of data for all teachers, thus requiring districts to employ a warehousing system. Programs are available for school districts to assist with the collection, organization, and dissemination of data required for teacher evaluations. OASYS is one program that maintains the data for teacher evaluation processes. The annual cost of OASYS for one school district in Central New York is $2,079 annually (B. Furlong, Personal Communication, April 28, 2015). This cost does not include the annual fee of $1,100 for the program maintenance fee. Furthermore, districts have added administrators to compensate for the additional evaluation time necessary under APPR. These financial costs are in addition to the cost of time away from the classroom and other administrative duties that are required with educating students.

Administrators have found their time spent on conducting multiple classroom evaluations, as required by the APPR law, increasing by as much as 50% (J. Regan, Personal Communication, April 27, 2015). More time is spent conducting classroom observations and writing up the evaluations than under previous evaluation guidelines (M. MacIntosh, Personal Communication, April 27, 2015). These intangible costs affect administrators’ abilities to complete and focus on requirements outside of teacher evaluations. While some districts have hired additional administrators to assist with
evaluations, others are not able to do so because their budgets cannot handle the additional costs.

Implementation of APPR did include a share of federal grant money to assist with the costs incurred with the process. Individual districts and NYSED shared the federal grant money from RTTT. Table 2.4 summarizes the parceling of district funds to New York State public schools.

Table 2.4

<table>
<thead>
<tr>
<th>New York State Public School RTTT Allocations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Lowest allocation</td>
</tr>
<tr>
<td>Highest allocation</td>
</tr>
<tr>
<td>Average allocation</td>
</tr>
<tr>
<td>Average allocation</td>
</tr>
</tbody>
</table>

Note. The big five city school districts comprise Albany, Buffalo, New York City, Rochester, and Syracuse. NYSED (2013a) website provides data regarding RTTT disbursements.

With such a large disparity in district funding, the RTTT application process allowed districts the option of pooling their resources to form network teams (OCM BOCES, 2011). The Central New York school districts chose to pool their resources and created a network team that consists of 19 public school districts. Syracuse City School District is part of the BOCES consortium in Central New York, but they chose not to pool resources with the network team.

Part of the task assigned to the network teams is to serve as a conduit between NYSED and the school districts (OCM BOCES, 2011). Monthly meetings offer training for lead evaluators and information sessions for superintendents, assistant
superintendents, and building-level school principal meetings. These sessions focus on
the implementation of the CCLS that is part of the RTTT grant as well as teacher
evaluations.

**Theoretical framework.** The theoretical framework for this research is
performance feedback theory. In 2003, Greve designed performance feedback theory,
and his work has received 5,262 citations since 2009 (Google Scholars, 2014). A
majority of these citations occurred during the economic downturn that hit the United
INSEAD School of Business. His research and focus on organizational innovation and
creativity led to the development of his performance feedback theory.

Performance feedback theory holds that individuals and organizations cyclically
review and compare their performance to previous history or comparable models as a
form of continuous improvement (Greve, 2003). Based upon this cyclical review,
decisions are made to either change practice or continue with current practice. The goal
of the individual or organization is to attain performance expectations.

Performance feedback theory provides the framework to review teacher
evaluations. Annual evaluation of teachers occurs with the goal of improving
performance in the classroom. This theory shares basic elements that are included in
teacher evaluation rubrics: (a) goal-setting, (b) performance feedback, and (c)
establishment of aspiration levels to change practice (Greve, 2003).

Greve (2003) is the key thinker behind performance feedback theory. His work
was heavily influenced by Cyert and March’s (1963) behavioral theory of the firm.
Developed in the early 1960s, behavioral theory of the firm focuses on decision making
and profit knowledge (Cyert & March, 1963). This foundational core is also present in performance feedback theory, although Greve provided a different approach to address decision making.

Steeped in organizational theories, the theory of performance feedback sets itself apart with a focus on how organizational performance influences decision making during times of uncertainty. Greve (2003) identified *slack, risk-taking, and aspiration levels* as critical components that induce change. Creating an atmosphere that allows individuals with an opportunity to design strategies to address problems is known as slack. Decentralized decision making is optimal for slack to achieve the most benefit. Most information to address and correct weakness is found in the lower tiers of an organization (Greve, 2003). Employing slack allows the organization with the ability to tap this resource in a creative and productive manner.

Organizations may guide decision making through their strategic use of creating slack situations. Time allowance for research and development is a form of slack. Organizing top performers and challenging them with a problem-solving dilemma is another form of slack (Greve, 2003).

In an educational environment, slack is found in common planning time, in-service opportunities, faculty meetings, and professional development days. Administrators have the ability to group teachers according to organizational needs. Greve (2003) found that once the atmosphere is crafted, the premise is that the people with the most working knowledge of the problem create ideas that address identified problems.
Originally developed in the behavioral theory of the firm, satisficing is highly relevant in performance feedback theory. Satisficing is the process of locating an alternative to reaching a goal, and it is grounded in the belief that decision makers have limited information, attention, and processing ability (Greve, 2003). Faced with limited information, decision makers are prone to make decisions based on goals. During the slack process, individuals develop alternative and, at times, competing ideas to address problems. This process is followed by an evaluation of alternatives until one alternative satisfies the goal that follows slack.

Goals and feedback accelerate learning (Greve, 2003). Performance improves when goal setting and feedback occur independently of one another. Greve (2003) found that people employed sophisticated problem-solving strategies in a focused effort to achieve clearly identified goals with performance feedback. Performance feedback theory addresses this aspect with the introduction of aspiration levels.

The central idea of performance feedback theory is the use of aspiration levels by decision makers to “evaluate organizational performance along an organizational goal dimension” (Greve, 2003, p. 41). Aspiration level is defined as “a reference point that is psychologically neutral” (Kameda & Davis, 1990, p. 56), and “the smallest outcome that would be deemed satisfactory by the decision maker” (Schneider, 1992, p. 1053). Aspiration levels serve as performance targets that are designed to be met, not exceeded (Greve, 2003). Changing performance for survival within an organization is an example of an outcome-based aspiration level.

Performance feedback affects behavior both in searching for the aspiration as well as the performance once it is established. Provision of feedback and implementation of
clear and specific goals by managers often result in the creation of aspiration levels that intuitively benefit an organization (Greve, 2003). Feedback (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2013) and goal setting (Darling-Hammond, 2013) evoke change in teacher practice.

Performance feedback systems include: (a) choosing goals, (b) generating aspiration levels, (c) finding solutions, (d) evaluating risk, and (e) making decisions (Greve, 2003). These performance feedback processes occur in educational settings. Of particular interest to the researcher’s study is the relationship of setting aspiration levels, evaluating risk, and making decisions as they apply to change in classroom instructional practice. APPR is predicated on the fact that change will occur because of that process. My study delves into what influence APPR has on teachers’ changing their practice.

Criticism of performance feedback theory. Behavioral theory of the firm is an alternate theory that addresses change in practice. Performance feedback theory is heavily steeped in behavioral theory of the firm. Decision making and goal setting are concepts of behavioral theory of the firm that Greve (2003) expounded upon with the creation of performance feedback theory. Greve proved that performance does enjoy continuous change, as it is predicated upon performance shortfalls. He furthered that an atmosphere conducive to development of alternatives for change to improve performance needs to be maintained by decision makers. Finally, Greve addressed what he viewed as a shortcoming in behavioral theory of the firm: How does an organization choose appropriate solutions for problems?

Performance feedback theory supports the study of teacher evaluation because it encompasses the nuances associated with education. Setting aspiration levels through
comparison with social groups, viewing performance levels through the lens of APPR composite score reports, and rubrics chosen for the evaluation process are aligned with key aspects denoted by Greve (2003) in performance feedback theory.

Evidence that performance feedback theory works. Two studies specifically employ performance feedback theory as the impetus for study. Greve (2003) found support in both the shipbuilding and radio studies for his theoretical hypothesis that innovation and change increase when organizational performance is low. The studies illustrated that when aspiration levels were met or exceeded, change in performance decreased. When aspiration levels were not met, change in performance increased.

Greve (2003) conducted a 26-year study of all major Japanese shipbuilders to find the effect of performance on research and development intensity. The study employed 11 firms with 230 collective years as its basis. These shipbuilding firms had modest research and development intensity at the time of the study, 1.4%, on average. The study controlled the number of employees, annual production, growth of shipping income, and oil freight rate, as these are industry-specific elements. The control variables were tested against performance aspiration levels and slack differentials. Absorbed slack and increased performance found a significant correlation at .055 when p < 01. The multiple correlation coefficient was significant with an R-square of .478.

Performance and slack are important elements for explaining research and development. Slack provides the impetus for research and development to occur. When performance is meeting or exceeding aspiration levels, slack is not prioritized, thus research and development is less robust. Research and development increased when competing shipbuilding companies brought new items to market and when internal
aspiration levels were not met. A linear relationship was apparent in the study when regressing performance, slack, and aspiration levels were the control variables (Greve, 2003). The initial study by Greve (1998) focused on changes in radio station formats, which is a niche industry. Unintentional sampling of 160 radio markets in the United States from January 1, 1984 through December 31, 1992 provided data for the study. Descriptive statistics were utilized, along with a mathematical equation, for determining aspiration levels. Format change was the dependent variable, and it included all changes, such as new format, innovative format, satellite, and other production change. There were 13 independent variables in the study, of which density, market share, station changes, and market changes were included.

When density of the market was strong, negligible changes occurred with marketing or within the station. This trend is evident in the data and supports the hypothesis that high performance on any measure reduces the likelihood of a format change (Greve, 1998).

Aspiration level determination is computed using a mathematical formula. Greve (1998) defined the symbols as A for aspiration level, P for performance, t for time period, and α for adjustment parameter. The formula for establishing an aspiration level is: \( A_t = \alpha P_{t-1} + (1- \alpha)A_{t-1} \). Performance feedback theory incorporates numerous mathematical equations because this theory applies to performance and production at the organizational level. The field of education does not offer the same hard concepts as that available in the business model, but the tenets of change are applicable.

Aspiration levels for teachers are found in the APPR issuance of a composite score report at the conclusion of the school year. This report includes both a numerical
and a categorical rating. These ratings are converted into performance aspiration levels that are crucial components of performance feedback theory.

Aspiration levels are determined based upon performance feedback or a comparison to similar models (Greve, 2003). For educators, environmental feedback may occur in the following forms, although it is not limited to them: (a) form of personnel evaluations, (b) student performance, (c) state mandates, (d) district initiatives, (e) community expectations, and (e) board of education goals.

Comparison tools can develop aspiration levels in education settings between teachers, departments, or buildings. Teachers may establish an aspiration level by comparing student achievement in a course from year to year. Aspiration levels may also be set based upon comparing APPR scores and categorically rating with peers of similar experience. Comparisons between individuals based upon similar traits is also found in social comparison theory (Greve, 2003). Performance feedback theory has elements of social comparison theory embedded in it.

Search and slack are indicative of a teacher’s decision-making process for determining how to respond to the internal and external performance feedback. Specific recommendations or comments from APPR may provide stimuli to address problems. Other sources of search may include attendance at conferences, professional learning communities, pedagogical books, journal articles, or faculty room discussions.

The last phase in the performance feedback theory is the risk phase. This is where teachers determine if change is necessary. Greve (2003) stated that, theoretically, when performance is at or above the aspiration level, there is little interest in change. Change
and the tolerance to risk change occurs when performance falls below the organizational or individual aspiration level.

In summary, performance feedback theory is a diagnostic tool used to discover problems (Ryan, 2004), and it is a cyclical process for continuous improvement (Greve, 2003). External and internal feedback is referenced against set goals and aspiration levels. When goals and aspiration levels are attained, decision makers are less inclined to take action to improve performance. Slack search is utilized to determine if a course of action is prudent given the performance feedback coupled with risk tolerance.

Conversely, when aspiration levels are not attained and a problem is identified, there is an increased tolerance to risk. Search and decision-making processes follow identification of a problem. Decision makers often establish goals as part of the solution to improve performance toward aspiration levels. According to Greve (2003), this process is cyclical as organizations and individuals engage in ongoing improvement.

**Change in individual practice.** Organizational change occurs as a discontinuous process that includes periods of instability and change interspersed with equilibrium (Greve, 2003; Marshall, 2011). Marshall (2011) identified three types of change: (a) sustaining, (b) disruptive, and (c) new market disruptive. Sustaining change does not disrupt current practice, because it invokes improvement to existing schema. An example of a sustaining change includes sending scanned images in an email instead of using a FAX machine. Disruptive change reshapes existing processes to improve performance. This type of change is problematic when current processes appear to be successful because it is a natural human tendency to protect current cultures and processes. Robotic replacement of assembly line workers is an example of a disruptive change. New-market
disruptive changes occur when existing schema are limiting and cannot work in the existing market. Earning a college degree online instead of attending a brick and mortar college is an example of a new-market disruptive change (Marshall, 2011). The overall change process tends to follow a cycle of instability that leads to change, and it is then followed by periods without change (Marshall, 2011; Schein, 1995).

Lewin’s change theory (1947) acknowledged three similar categories for change, although he referenced them as unfreeze, moving, and refreeze (Schein, 1995). When dissatisfaction occurs in an organization, unfreeze occurs. Data depicting failed organizational expectations creates dissatisfaction or frustration and this starts the change process. Becoming unfrozen allows individuals or organizations to reframe their current mindset, thus allowing a new perspective to view the problem, and then they act accordingly. The symmetry of reframing and the acceptance of the disconfirming date creates the motivation to change. The act of changing is termed moving, and when equilibrium is ascertained, refreezing occurs. Refreezing is critical as this stage because it allows the change to become stable. Schein (1995) reported that failure to reach personal or organization goals results in change, and this process reaffirms Greve’s (2003) performance feedback theory.

The overall goal of the individual or organization is to attain and maintain performance expectations (Greve, 2003). Greve (2003) theorized that when performance is at or above the aspiration level, there is little interest in change. The central idea of performance feedback theory is the use of aspiration levels by decision makers to “evaluate organizational performance along an organizational goal dimension” (Greve, 2003, p. 41). An aspiration level is defined as “a reference point that is psychologically
neutral” (Kameda & Davis, 1990, p. 56), and “the smallest outcome that would be deemed satisfactory by the decision maker” (Schneider, 1992, p. 1053). Aspiration levels serve as performance targets that are designed to be met, not exceeded (Greve, 2003). Change, and the tolerance to risk change, occurs when performance falls below the organizational or individual expectancy level. Marshall (2011) identified this as disruptive, and Lewin’s change theory referred to this element of change as moving (Schein, 1995). The impetus for change begins with dissatisfaction or failure to reach expected performance levels (Greve, 2003; Schein, 1999).

Fullan and Steigebauer (1991) described change as a highly personal process that involves a multitude of feelings and emotions, one of which is anxiety. Focusing on change in the field of education, they furthered that due to the cellular structure of schools, teachers spend most of their time in isolation. Moving to a more collegial working relationship, which fosters open communication and support, leads to a higher likelihood of successful implementation of change. Fullan and Miles’s (1992) seven propositions for successful change include: (a) change requires making personal meaning, and is therefore a learning process; (b) change is not defined, it is fluid and requires continuous adjustment; (c) problems are to be expected during the change process, and they are helpful; (d) change requires continuous support, and includes increased costs; (e) hands-on work is required, so change cannot be managed from afar, it requires leadership on site; (f) change is systemic and therefore contains many nuances—all of which need to be learned and understood; and (g) large-scale change is not successful if a bureaucratic model is used, because it requires local leadership. Fullan (2006) later surmised that motivation is the one word that describes the key facet for
change to occur. Unless people are motivated, each one—individually—to put forth the
effort and exertion to change, it will not occur (Fullan, 2006; Terhart, 2013).

Fullan and Miles (1992) and Terhart (2013) acknowledged that there are
fundamental reasons why change does not occur and why there is resistance to change in
the field of education. Misunderstanding and confusion about the change process, as it
affects the individual classroom teacher and the overall educational process, is one of the
more prominent reasons why educational change fails (Fullan & Miles, 1992; Terhart,
2013). When existing cultures and paradigms are replaced with practices that are not yet
known nor established, uncertainty and anxiety among teachers is common (Terhart,
2013). Fullan and Miles (1992) explained that this is a natural response to change, and it
should not be viewed as resistance. Human beings need to make sense of change, as it
personally relates to them, before they enact any change. Similarly, teachers’ beliefs and
attitudes are key components that determine whether they will implement and maintain
change in their instructional practice (Boardman & Woodruff, 2004). Knowledge about
the change process needs to be clearly explained and communicated by those seeking the
change (Fullan & Miles, 1992). Without ongoing support, which occurs by helping
teachers make meaning of the change, the process or program will not survive.

Performance assessment in the medical field. Human capital is the most
critical resource that needs to be tracked and focused on in the medical profession
(Emami & Doolen, 2015). As in the field of education, performance assessment in the
medical field is a complex priority. Effective implementation of performance
assessment, according to Collett and Starke (2008), continues to be a challenge. The
focus of assessment is to provide “opportunities for constructive, thoughtful feedback and
to identify future learning objectives” (Collett & Starke, 2008, p. 401). Implementation of this type of assessment requires a considerable commitment of time by both the assessor and the medical professional under review.

Part of the complexity lies in the varied views held by stakeholders regarding performance assessments. Views include, but are not limited to, encouraging excellence, employment of best practice, competence, fulfillment of statutory requirements, and distinguishing between medical practitioners for purposes of training and promotion (Collett & Starke, 2008). In order to address these expectations, assessment tools need to include a clear description of what is being assessed and the overall purpose of the assessment. Incorporating standards into the assessment tool allows for clear communication of expectations and goals. The Accreditation Council for Graduate Medical Education and the American Board of Medical Specialties outlined six domains of competence for assessment of medical practitioners (Collett & Starke, 2008).

Whereas assessment is required and part of good practice to further advancements in the field of medicine, it does incur financial investment. Performance assessment expenses include managerial time, technological programs and services, training for the assessors, and the time billed by the assessing physicians and the assessed physicians.

Performance assessment tools for the medical profession generally include use of volunteers who serve as pseudo patients, video observation of actual on-the-job performance, direct observations including technical procedures, peer assessment in the form of 360-degree feedback, patient questionnaires, and audits of medical records (Collett & Starke, 2008). These assessment tools include both qualitative and
quantitative data for the assessors. The direct observation modality is known as Workplace Based Assessment (WPBA) (Guraya, 2015).

WPBA combines direct observation in the workplace with specific and reflective feedback, thus combining elements of performance feedback theory. Guraya (2015) stated that WPBA is a formative assessment tool that promotes learning and is not a summative assessment tool. Observation of clinical performance, discussion of clinical cases, and feedback from peers, coworkers, and patients are the main components of WPBA (Guraya, 2015). Feedback is delivered using the 360-degree format where peers complete an assessment, team assessment of behaviors occurs, and a patient satisfaction survey is employed.

Due to the complexity of performance assessments, along with the time required for adequate implementation, the medical profession should prioritize resources and employ a narrow focus on areas of performance to achieve success (Emami & Doolen, 2015). Prioritizing and closely tracking fewer metrics is less costly and strategic. Emami and Doolen (2015) stated that ongoing employee development is synonymous with providing high-quality care. Employee development is part of a performance assessment program.

**Chapter Summary**

Efforts to improve student achievement in the United States are currently underway with an increased focus on holding teachers accountable for student learning. State laws now require the implementation of evaluation systems to ensure effective teachers are in the classrooms. New York State passed Education Law §3012-c to earn a competitive federal grant known as RTTT. The RTTT grant provided New York State
with just under $700 million that was earmarked for use in implementing a rigorous teacher evaluation system with an overall goal of improving student achievement. Consequently, NYSED implemented an APPR system designed to effectively evaluate teachers, while simultaneously addressing the achievement gap for students.

Researchers have documented that performance appraisals have the potential to improve teaching practice. Procedural facets of the evaluation process serve as a determinant as to whether teachers will implement feedback garnered from performance evaluations. Trust in the process between administrators and teachers, and trust in the purpose of the evaluation data, is critical for creating sustained and systemic change in instructional practice.

District professional development is guided by administrative and teacher feedback from the evaluation process. Performance appraisals serve two purposes in education because they provide guidance for professional development and changes in the instructional practice in the classroom. When designed with the intent to inform practice, standards-based evaluation procedures lessen forms of bias and improve teacher self-efficacy. Given the premise that the revised APPR system will improve education, it is imperative that information gleaned from ongoing performance appraisal studies be implemented with fidelity, in a collaborative process, with all classroom instructional practices to improve student performance.
Chapter 3: Research Design Methodology

Introduction

One standard deviation increase in teacher effectiveness has a value of approximately $330,000 to $760,000 when factored into a student’s lifelong earning potential (Staiger & Rockoff, 2010). Kersten and Israel (2005) stated that the national focus on accountability in education occurred with the publication of *A Nation at Risk* in 1983. Accountability for student achievement remains a national focus with continued congressional reauthorization of the Elementary and Secondary Education Act (ESEA). Prior to 2010, teacher evaluation in New York State was designed and implemented at the school district level with protocols and procedures varying in number of annual observations, inclusion of student data, and training of evaluators.

With congressional reauthorization of the ESEA in 2009, President Obama established a goal to have a great teacher in every classroom and a great principal in every school (USDE, 2010). The Race to the Top fund was subsequently established to achieve this goal by providing competitive grants to states that enacted ambitious and comprehensive reforms that focused specific attention on teacher accountability.

Passage of Education Law §3012-c and subsequent §3012-d in New York State impacts annual professional performance reviews for teachers by requiring annual evaluations. APPR also requires a final quality rating for all teachers based upon their evaluations and student performance. Law (2007) reminded that “evaluations are powerful tools that can inhibit or empower employees in relation to quality and
productivity” (p. 19). Passage of NYS Education Law §3012-c occurred for two main reasons: first, for the state to acquire close to $700 million dollars in RTTT funding from the federal government, and second, to change the instructional practice of classroom teachers to improve student achievement. Therefore, it is imperative for evaluation tools to do more than identify strong and weak teachers (Papay, 2012). To be effective, teacher evaluation tools must assist teachers in assessing their performance, as well as provide a road map for developing pedagogical skills that will improve achievement for each student. Teacher evaluation theoretically provides an impetus for teachers to change their classroom instructional practice to improve student achievement and drive professional development for school districts.

School districts have spent numerous hours negotiating the implementation of the APPR plans by dedicating professional development days to APPR training. Administrators spend days out of the district receiving training so that they can conduct standards-based evaluations, and student test scores have become the focus of the educational arena. Given the significant investment of time, money, and precious school resources, there is an expectation that teacher evaluations under APPR will result in change of classroom instructional practice that, in turn, will result in improved student achievement.

The research design for the study is an exploratory, sequential, mixed-methods study. Operationalization of this two-phased study is shown in Figure 3.1. Exploratory mixed methods use qualitative data to inform the process of data collection in a quantitative study (Creswell, 2014).
Mixed-methods research blends information from both qualitative and quantitative research (Creswell, 2003). Sequential implementation of qualitative and quantitative research confirms information from different types of data and allows for a deeper understanding of the phenomena. In an exploratory strategy, the collection of qualitative data identifies emergent themes and findings that then guide the quantitative data collection process. An exploratory strategy encompassing two phases were utilized for this study.

Phase one employed the use of two qualitative focus groups to address the first two research questions:

1. From the perspective of selected, experienced lead evaluators within Central New York school districts, who have 2 years of experience under the new APPR process, and what are the critical components of classroom instructional practice contained in the state-approved evaluation rubrics?

2. From the perspective of selected teachers in Grades 9-12 within Central New York school districts, who have 2 years of experience under the new APPR process, what are the primary impetuses for teachers to change classroom instructional practice?
Employing a qualitative approach allows for a detailed understanding of the emergent issues and empowers stakeholders to share their experiences. Lead evaluators with at least 2 years of experience evaluating teachers under the newly revised APPR process identified the critical components of classroom instructional practice. Teachers in Grades 9-12, who have 2 years of experience under the new APPR process within Central New York school districts, identified the primary impetuses for teachers to change classroom instructional practice.

Creswell (2013) stated that human interactions are difficult to statistically measure, and they overlook the importance of the individual in the process. Consequently, exploring the critical components of classroom instructional practices through qualitative methods enables certified lead evaluators with an outlet for their voices that a quantitative approach would not be able to assess. Addressing the importance of the individual’s voice in a focus group (Creswell, 2013), the researcher facilitated the teachers’ discussions of the impetus for change in instructional practices.

Once the critical components of classroom instructional practices and impetuses for changing these practices were identified, phase two commenced with a quantitative study. A survey explored in what components teachers made the most change in instructional practice during the 2013-2014 and 2014-2015. Additionally, the surveys identified what were the primary impetuses for these changes. Creswell (2014) stated that surveys provide a “numeric description of trends, attitudes, or opinions of a population by studying a sample of that population” (p. 155). The design, distribution, and disaggregation of the survey for the study used the Qualtrics online survey program. Descriptive statistics analyzed the findings from the survey. Based upon the need to
investigate the critical components of instructional practice and the impetus to change these practices prior to conducting the quantitative aspect of the study, an exploratory sequential mixed methods methodology was the optimum choice for this research study.

The study examined the extent to which public school teachers of English, math, science, and social studies in Grades 9-12 are changing classroom practice based upon the implementation of New York State Education Law §3012-c (APPR). The collected data answered the following questions.

1. From the perspective of selected, experienced lead evaluators within Central New York school districts, who have 2 years of experience under the new APPR process, and what are the critical components of classroom instructional practice contained in the state-approved evaluation rubrics?

2. From the perspective of selected teachers in Grades 9-12 within Central New York school districts, who have 2 years of experience under the new APPR process, what are the primary impetuses for teachers to change classroom instructional practice?

3. Given the critical components of classroom practice, as identified by Q1, from the perspective of selected English, math, science, and social studies teachers in Grades 9-12 within Central New York school districts, in which component(s) did they make the most changes in instructional practice during the 2013-2014 and 2014-2015 school years?

4. Given the changes identified by the same participants in question 3, what was the primary impetus for changes made in their classroom instructional practice?
**Research Context**

The context for the study focused on public school districts within the Central New York region as identified in Table 3.1. Even though Central New York comprises Oswego, Cayuga, Onondaga, Cortland, and Madison counties, the study only included participants from Cortland, Onondaga, and Oswego counties.

Table 3.1

*Public School Districts Within the Central New York Region*

<table>
<thead>
<tr>
<th>County</th>
<th>Number of Public School Districts</th>
<th>Student Population</th>
<th>Full-Time Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oswego</td>
<td>9</td>
<td>23,622</td>
<td>1,777</td>
</tr>
<tr>
<td>Cayuga</td>
<td>7</td>
<td>11,169</td>
<td>873</td>
</tr>
<tr>
<td>Onondaga</td>
<td>18</td>
<td>76,861</td>
<td>5,664</td>
</tr>
<tr>
<td>Cortland</td>
<td>5</td>
<td>7,248</td>
<td>585</td>
</tr>
<tr>
<td>Madison</td>
<td>10</td>
<td>11,664</td>
<td>890</td>
</tr>
</tbody>
</table>

*Note.* The Central New York Region comprises five counties. Demographic data provided by an Internet data project combining government and other sources for New York Schools (www.NewYorkSchools.com).

Phase one of the study included two focus groups: one with selected and recruited lead evaluators, and the second with selected and recruited classroom teachers. Lead evaluators from a combination of 24 rural, suburban, and urban public school districts within Cortland, Onondaga, and Oswego counties were invited to participate in a focus group during the first phase of the study. The lead evaluators identified critical components of classroom instructional practice contained in the state-approved evaluation rubrics.
A convenience sample of recruited and selected teachers from Onondaga County participated in a second focus group during the first phase of the study. This focus group identified primary impetuses for teachers to change classroom instructional practice.

During the second phase of the study, teachers from a combination of suburban and urban public school districts within Cortland, Onondaga, and Oswego counties received a link to the survey. Data collected from the survey identified component(s) where teachers made the most changes in instructional practice over the previous two school years, 2013-2014 and 2014-2015, and the impetus for the change(s).

Central New York enjoys a stable population, with females slightly outnumbering males by 51 to 49% (Vink, 2011). The region is predominantly White, non-Hispanic, with trending data on ethnicity remaining stable. White and non-Hispanic is claimed as an identifying ethnicity by 88%, with Black or African American anchoring the second highest ethnic group with 7% of the population. White-Hispanic, Asian, and those individuals identifying with two or more ethnic groups make up the remaining 5% of the ethnic breakdown of Central New York.

NYSED publishes an Accountability Status of District and School Report annually. Schools earn the labels of Local Assistance Plan, Focus, or Good Standing on their school district report cards based upon student achievement and graduation rates. Twenty-three school districts in the study were identified as being in Good Standing, and one district was identified as a Local Assistance Plan (NYSED, 2014a). High Performing Rewards Schools meet the following criteria: (a) they make annual yearly progress in English, math, science, and with their graduation rate; (b) they show bottom quartile growth for elementary and middle school students; (c) their graduation rate
exceeds 80%; and (d) the schools exceed the state average for student graduation with either a Regents diploma with an advanced designation or a Career and Technical Education (CTE) endorsement. For the 2014-2015 school year, 25% of the school districts participating in the study had one or more schools within their districts that earned the High Performing Rewards School designation.

New York State has a list of 10 approved teacher evaluation rubrics posted on the EngageNY website from which districts may choose for teacher evaluations under APPR. The NYSUT rubric was chosen by 58% of the districts, with the Danielson model chosen by 29% (NYSED, 2014b). OCM BOCES has chosen to focus lead evaluator training on these two rubric systems based upon the districts’ preference for them.

**Research Participants**

Convenience sampling identified participants for phase one of the study, and phase two employed a single-stage sampling. Phase one consisted of two focus groups of selected and recruited participants. The first focus group that identified critical components of classroom practice included lead evaluators, and the next focus group that identified impetuses for changing instructional practice consisted of teachers. This phase provided qualitative data for the creation of a quantitative survey for data collection in phase two.

Data from the qualitative design of phase one informed the development of a quantitative survey for phase two. A single-stage sampling of teachers from New York public school districts in Cortland, Onondaga, and Oswego counties provided the participants for phase two. An Internet-based survey collected the information to address
what changes occurred in instructional practice and to what extent APPR served as the impetus for change.

**First qualitative focus group.** Recruited and selected lead evaluators participated in a focus group to identify critical components of classroom instructional practice contained in the state-approved evaluation rubrics. Lead evaluators, district personnel who conduct teacher evaluations, with at least 2 years of experience evaluating teachers under the §3012-c APPR law, were selected and recruited to participate in the focus group. Selection of lead evaluators assured their evaluation experience with teachers of English, math, science, and social studies in Grades 9-12. An email (Appendix A) initiated participation of the lead evaluators to participate in the focus group. An introductory letter was then sent to lead evaluators participating in the study (Appendix B).

Selected lead evaluators received an informed consent form (Appendix C) required for participation in the study at the commencement of the focus group. Receipt of the signed informed consent form was required for the participation. Participants were informed that data collected from the focus group activity would inform the construction of a survey for teachers in phase two of the study. Clearly noted on the consent form was a description of the study and guarantee of confidentiality. Numbers assigned to participants during the transcribing and coding process guaranteed confidentiality. They were also informed that the researcher is a student in the Doctorate in Executive Leadership Program at St. John Fisher College.

**Second qualitative focus group.** Recruited and selected teachers participated in a focus group to identify impetuses for changing classroom instructional practice. Using
a convenience selection, teachers of Grades 9-12 were recruited for the study from two public school districts in Central New York. Recruitment included tenured teachers of English, math, science, and social studies. An email (Appendix D) personally inviting selected teachers to participate in the focus group initiated participation. Representation in this phase encompassed teachers who had experience with being evaluated under the §3012-c APPR law during the 2013-2014 and 2014-2015 school years. An introductory letter was then sent to teachers participating in the study (Appendix E).

These teachers received an informed consent form (Appendix C) required for participation in the study at the commencement of the focus group. Receipt of the signed informed consent form was required for the participants to engage in the focus group. Clearly noted on the consent form is a description of the study and guarantee of confidentiality. Numbers assigned to the participants during the focus group and transcribing process guaranteed confidentiality. Participants were informed that the data collected from the focus group activity would inform the construction of a survey to be sent to teachers in other districts for the second part of the study. They were also informed that the researcher is a student in the Doctorate in Executive Leadership Program at St. John Fisher College.

**Phase two.** The researcher directly contacted public school districts in Cortland, Onondaga, and Oswego counties to obtain permission for teacher participation in the quantitative phase of the study (Appendix F). Initial requests occurred through emails, followed by an email electronically confirming district permission.

**Quantitative survey.** Phase two involved the administration of a quantitative survey to teachers of English, math, science, and social studies in Grades 9-12 with at
least 2 years of evaluation experience under the revised APPR process. District permission for teacher participation was obtained through email correspondence.

School districts agreeing to participate in the study received a link to the Qualtrics designed survey (Appendix G). In turn, district personnel emailed the link to their entire faculty. Qualtrics employs skip logic that allowed all teachers to begin the study, but only those meeting the criteria of the study were able to complete the entire survey.

The teachers provided informed consent by checking the appropriate box located on the consent form prior to the introduction to the survey. This check box confirmed that the participant: (a) was aware of the parameters of the study, (b) understood that participation is voluntary, and (c) guaranteed that their participation in the study was anonymous. Anonymity was guaranteed in that the researcher never directly emailed or contacted the participants. Personnel from the districts consenting to participate emailed a link to the survey to all faculty members. Every teacher in the district had the opportunity to participate in the study.

Recruited and selected teachers from the researcher’s school district, who were not participating in the study, field tested the survey during August 2015. Field tests included testing the skip-logic flow of the survey and the validity of the questions.

**Instruments Used in Data Collection**

Mixed-methods research blends information from both qualitative and quantitative research (Creswell, 2003). Sequential implementation of qualitative and quantitative research confirms information from different types of data and allows for a deeper understanding of the phenomena. In an exploratory strategy, the collection of qualitative data identifies emergent ideas and findings that then guide the quantitative
data collection process. The study encompasses an exploratory strategy, which is divided into two phases.

Phase one commenced with a qualitative approach to identify criteria on the New York State APPR rubrics that reflect critical components of classroom instructional practice. A qualitative approach allows for a detailed understanding of the emergent issues and empowers stakeholders to share their experiences. The study employed focus groups for the qualitative phase. Creswell (2013) stated that human interactions are difficult to measure statistically and overlook the importance of the individual in the process. Consequently, exploring the performance evaluation and classroom instructional practices through focus groups captured certified lead evaluators’ voices that would not be possible with a quantitative approach. Efforts to obtain a representative sample entailed selecting and recruiting certified lead evaluators to participate in the study.

Data collection for the study occurred in two stages, beginning with two focus groups addressing the critical components of classroom instruction and impetuses that evoke change in instructional practice by classroom teachers. Stringer (2014) described a focus group as synonymous with a group interview. Rossi, Lipsey, and Freeman (2004) added that focus groups allow stakeholders with an opportunity to have in-depth discussions on a particular topic. The strength of a focus group is that the certification training for lead evaluators provided symmetry in describing the critical components of classroom practice contained in the state-approved APPR rubrics. This symmetry offset the small participation number of the convenience-sample method.
**Teacher focus groups.** Teachers with classroom experience and with at least 2 years of evaluation experience under the revised APPR process provided genuine data regarding the impetus for changing instructional practices. Discussing a series of semi-structured questions, all members had opportunities to share experiences and beliefs during the focus group (Appendix H). The focus group included six participants to ensure each participant’s ability to engage in the dialogue (Creswell 2013). The focus group session lasted 55 minutes and was digitally recorded. Fowler (2014) stated that it is imperative that participants understand how the researcher will utilize information from the interview, and that the data obtained is helpful. In addressing this recommendation, the participants received a description of the study one week before the focus-group meeting. The participants were assigned a number for recording, transcribing, and coding thus allowing for participant confidentiality. All field notes, coding journal entries, and data is locked in a filing cabinet at the researcher’s home and will be destroyed 3 years after publication of the research.

**Lead evaluator focus group.** An email to lead evaluators inviting them participate in the focus group initiated participation (Appendix A). Five selected and confirmed lead evaluators received an informed consent form (Appendix C), which was required for participation in the study prior to the commencement of the focus group. Receipt of the signed informed consent form was required for the participation. Clearly noted on the consent form was a description of the study and guarantee of confidentiality. Numbers assigned to the participants during the focus group meeting and transcribing process guaranteed confidentiality.
The focus group discussed a series of semi-structured questions, allowing all members opportunities to share experiences and beliefs (Appendix I). The focus group included five participants to ensure each participant’s ability to engage in the dialogue (Creswell 2013). The focus group session lasted 65 minutes and was digitally recorded. Fowler (2014) stated that it is imperative that the participants understand how the researcher will utilize information from the interview, and that the data obtained is helpful. In addressing this recommendation, the participants received a description of the study one week before the focus group meeting. The participants were assigned a number for recording, transcribing, and coding thus allowing for participant confidentiality. All field notes, coding journal entries, and data are locked in a filing cabinet at the researcher’s home and will be destroyed 3 years after publication of the research.

Data collected from the focus groups in phase one determined the content of the survey for data collection in phase two. Phase two included the implementation of quantitative methodology to answer research questions 3 and 4 to address what changes have occurred in classroom instructional practice along with the impetus for the change.

Survey development implemented the use of closed questions. Meaning from closed-question responses are more reliable, and when administered electronically, offer an easier format for participants to complete (Fowler, 2014). The Qualtrics online survey tool allows for designing instruments that minimize clutter, employs skip logic negating the need for respondents to answer specific questions, employs forced choice so that respondents are not able to skip questions, and allows for a clean organization that is
appealing to the eye. Fowler (2014) identified the attributes of the computer-assisted instruments as critical for ensuring an easy way to use the questionnaire.

The first focus group identified critical components of instructional practice based upon NYSUT and Danielson APPR rubrics. The teachers engaged in a discussion of impetuses for changing classroom instructional practice during the second focus group. Creation of the survey for phase two employed data collected from both focus groups. Phase two identified the areas of classroom instructional practice that have changed over the previous 2 years and the impetuses for the changes for ELA, math, science, and social studies teachers of Grades 9-12.

Field testing the survey occurred with a convenience sample of 40 teachers. Fowler (2014) noted that “a particularly important function is to test the usability of the instrument, both the questions and the layout from the interviewers’ perspectives” (p. 106). Field testing also enables the assessment of the survey’s validity. When surveys accurately assess the premise of a study with little margin for error, they are deemed valid (Fowler, 2014). Face validity occurs when experts in the field are able to ascertain that the survey is measuring the intended topic by reviewing the survey questions. By having an array of teachers field test the survey, the researcher validated that the survey would accurately filter participants and guarantee that only ELA, math, social studies, and science teachers of Grades 9-12 completed the survey.

Thirty-eight teachers field tested the survey over one week, followed by two debriefing sessions. The recruited teachers reviewed the survey for clarity and relevance to the research questions to enhance the validity of the survey. The debriefing sessions included review of the questions to determine what the participants felt was the focus and
intent and to estimate the time required to complete the survey. The goal was to create a survey that took 15 minutes to complete. Comments from the debriefing sessions informed the finalization of the survey for phase two of the study.

Data collection from an online survey using Qualtrics was the basis of phase two of the study (Appendix G). In this phase, teachers from Cortland, Onondaga, and Oswego public school districts, where superintendents granted approval for participation, received the survey link (Appendix J). Reminder emails were sent 10 and 12 days after the initial email to remind teachers of the opportunity to participate in the survey.

Inclusive in all communication throughout the process was the acknowledgement that the data would be aggregated, thus ensuring participant anonymity. Survey access by teachers closed after 14 days.

The data obtained from the focus groups in phase one determined the content and development of the survey for phase two. The intention was to create an Internet-based survey that took less than 15 minutes to complete. Internet-based formats allow for an ease of design that employ skip logic, can employ information from previous questions to address the sequencing of the survey, and can identify inconsistencies in responses at the point of entry (Fowler, 2014). Additionally, Fowler (2014) noted that Internet-based surveys are cost-effective, have the ability to return data quickly, and include the benefits of self-administered surveys. The drawbacks to this Internet-based survey was that it required district approval prior to the administration and the enlistment of a district employee to email the survey link to the faculty.

The survey commenced with demographic data questions to confirm criteria for completion of the survey. Following this section on the survey, teachers were asked to
identify areas of classroom instructional practice where they have made changes over the previous 2 years. For each area of classroom instructional change identified, teachers indicated the degree of change using tiered ranking. None, moderate, and significant were the three ranking levels from which teachers chose their degree of change. Finally, the teachers also identified the impetus for each change of instructional practice from a list of pre-determined impetuses.

Data analysis from the survey employed the use of descriptive statistics from the Qualtrics Internet-based software. Analysis of the extent to which change is taking place, in what practices, and what precipitated the change was determined using descriptive statistics of mean, standard deviation, and range.

Response bias could not be determined by participation methods because participation was anonymous. However, emails were sent to district officials at the start of the two-week survey period and, again, seven days into the data-collection period. District officials emailed the link to the survey to district faculty. Sending the surveys to at least 10 districts found a return rate of 200 respondents based upon a 50% response rate. Fourteen school districts agreed to participate in the study.

**Data Analysis**

Creswell (2014) advised understanding the bias that is inherent with the researcher when designing a study. In an effort to mitigate potential bias, the study on APPR was conducted outside of the school district where the researcher works.

Creswell (2013) defined coding as “aggregating the text or visual data into small categories of information, seeking evidence for the code from different databases being used in a study, and then assigning a label to the code” (p. 184). Saldana (2013) mirrored
the definition of coding as a word or phrase that captures the essence of the visual or verbal data. Creswell further recommended starting with a short list of tentative codes to begin the process, and then expand as warranted. Following the coding of the data, categorical themes were developed to organize emergent information. The codes took the form of current research terms or terms developed by the researcher.

Qualitative research involves identification of emergent data through the coding process (Creswell, 2013). Fowler (2014) advised carefully checking and rechecking throughout the coding process to lessen the possibility of errors. Saldana (2013) referred to this practice as first-cycle and second-cycle coding, although the larger the database, the greater the likelihood of further coding cycles.

Employing the manual coding process identified by Creswell (2013) and Saldana (2013), the study utilized codes developed by the researcher and included two cycles of coding. The cyclical nature of coding allows emergent phenomena to become categories for organizing data. Patterns begin to emerge as the categorization or codifying process takes place. Rules, identities, emotional aspects, hierarchical aspects, inequalities, causes, and effects are the types of data that Saldana (2013) denoted as the basis for coding. Perceptions, feelings, and tangible elements provided by participants in the focus groups were also coded, then codified, into categories and subcategories.

**Focus group one.** The first phase of the study employed a focus group to provide data on critical components of classroom practice that are found in the APPR rubrics used in the school districts in the study by certified lead evaluators. Fowler (2014) suggested recording open responses, verbatim, for coding purposes to minimize decision making while conducting the interview. Audio recording of the focus group discussion occurred
to ensure an accurate representation of the data. Due to the transparency and open nature of coding and data reduction, this process has a low potential for bias.

During the first manual coding cycle, the researcher looked for phrases or terminology that addresses the guiding questions from the focus group. Identification of patterns and emergent themes from the initial coding cycle occurs during the second and more in-depth cycle of the coding process. During this stage, attention centered on components and behaviors of instructional practice that evaluators can see when evaluating a teacher. In-depth coding allows for the identification of critical components of classroom instructional practice, the behaviors that represent these components, and the possibility of the ability of the evaluators to assess them. The researcher employed Saldana’s (2013) recommendation to maintain a codebook that identifies the codes and content descriptions.

**Focus group two.** The first phase of the study employed a second focus group to provide data on the impetuses for changing classroom practice by certified teachers of English, math, science, and social studies in Grades 9-12. An audio recording of the focus group discussion was conducted to ensure an accurate representation of the data. Due to the transparency and open nature of coding and data reduction, this process has a low potential for bias. Fowler (2014) suggested recording open responses, verbatim, for coding purposes to minimize decision making while conducting the interviews.

During the first manual coding cycle, the researcher looked for phrases or terminology that addressed the semi-structured questions from the focus group. Identification of patterns and emergent themes from the initial coding cycle occurred during the second and more in-depth cycle of the coding process. During this stage,
attention was centered on emotions or feelings that spurred teachers to change instructional practice in their classrooms. In-depth coding allowed for the identification of the rationale behind the impetuses to change practice, the behaviors that resulted in change, and the sustainability of these changes. The researcher employed Saldana’s (2013) recommendation to maintain a codebook that identified the codes and content descriptions.

**Quantitative survey.** The coded data was utilized to create a 10-question survey for phase two of the study. The Qualtrics Internet-based quantitative survey tool was utilized to design, distribute, and aggregate data during the second phase of the study. To collect data, the survey employed three-point ranking questions and closed questions that required forced responses. Analyzing and displaying the data to answer the research questions focused on the components of classroom instructional practice that changed over the previous 2 years. In addition, the analysis and display of the impetus for the described changes employed descriptive statistics of mean, standard deviation, and range.

The demographic data confirmed the selection of the participants to allow them to complete the survey. Teachers of English, math, science, and social studies in Grades 9-12 completed the survey. The participants selected from two main areas that described their classroom instructional changes. They then indicated on a three-tiered ranking scale the degree of change that occurred. The tiered ranks included *none, moderate*, and *significant*. The teachers chose a description from a list of impetuses for each classroom instructional change they instituted. The Qualtrics Internet-based survey tool allows skip logic to guide the participants through the survey, tailoring the sequence of questioning based upon answers provided.
The data analysis includes charts and graphs that identify areas of classroom instructional changes that were employed by the teachers over the previous 2 years (Table 4.1). The representation of the changes includes charts denoting the percentage of teachers who changed their practice in these areas. The impetuses, from the most significant change to the least significant change, are presented in a hierarchical figure in Figure 4.5.

Additionally, the data analysis denotes the impetuses for each change. A representation of the changes includes figures/tables denoting the percentage of teachers who made classroom instructional change based upon specific impetuses (Tables 4.2, 4.3, 4.4, 4.5, and Figures 4.5 and 4.6).

Summary

Teacher evaluation has become the platform for holding teachers accountable for student learning. A teacher’s evaluation now has the ability to empower or disenfranchise teaching quality and productivity (Law, 2007). APPR was revised in an effort to improve student achievement by way of spurring changes in classroom instructional practices. Feedback and professional development from the implementation of the revised APPR law in New York State should have resulted in improved classroom instructional practices. However, implementation of such standards-based evaluation systems has proven to be challenging as teaching is a complex process, with pedagogical aspects that are difficult to assess consistently, fairly, and accurately (Bereens, 2000). Consequently, this sequential mixed-methods research studied the extent to which APPR is changing classroom instructional practice for English, math, science, and social studies teachers in Grades 9-12 in Central New York.
Chapter 4: Results

Introduction and Research Questions

The study examined the extent to which public school teachers of English, math, science, and social studies in Grades 9-12 are changing classroom practice based upon the implementation of New York State Education Law §3012-d. Performance feedback theory posits that cyclical evaluation of performance causes change in practice that leads to improvement in outcomes. Given the considerable amount of time, effort, and financial expenditure on the part of school districts implementing APPR, in the hopes of changing teacher practice to improve student learning, it is critical to understand the extent to which APPR is changing instructional practice in the classroom.

Collected data answered the following research questions.

1. From the perspective of selected, experienced lead evaluators within Central New York school districts, who have 2 years of experience under the new APPR process, what are the critical components of classroom instructional practice contained in the state-approved evaluation rubrics?

2. From the perspective of selected teachers, in Grades 9-12 within Central New York school districts, who have 2 years of experience under the new APPR process, what are the primary impetuses for teachers to change classroom instructional practice?

3. Given the critical components of classroom practice, as identified by question 1, from the perspective of selected English, math, science, and social studies
teachers in Grades 9-12 within Central New York school districts, in which component(s) did they make the most changes in instructional practice over the past 24 months?

4. Given the changes identified by the same participants in question 3, what was the primary impetus for changes made in their classroom instructional practice?

A sequential mixed-methods research paradigm divided into two phases was employed for this study. The qualitative design of phase one answered the first two research questions and informed the quantitative second phase of the study that answered research questions three and four. Permission for teacher participation was obtained from public school districts in Cortland, Onondaga, and Oswego counties.

Phase one began with purposeful selection and recruitment of lead evaluators and teachers from the Central New York area to obtain a representative sample for two focus groups. Each focus group was designed to answer a specific research question. Lead evaluators identified critical components of classroom instructional strategies, thus answering research question 1. Teachers identified impetuses for changing classroom instructional practice, thus answering research question 2.

Selected and recruited teachers and administrators received phone calls and emails inviting them to participate in a focus group as part of a doctoral study. Five lead evaluators agreed to participate in the focus group identifying critical components of classroom instructional practice, and six teachers agreed to participate in a focus group identifying impetuses for changing classroom instructional practice. An email was sent confirming the teachers’ focus group for July 27, 2015 at 11:00 am, and a corollary email
was sent to the lead evaluators confirming their focus group for July 27, 2015 at 3:00 pm. Informed consent forms were signed at the beginning of each focus group meeting by all participants.

Data from the focus groups in phase one informed development of a survey for use in phase two of the study. The survey was designed to identify components in which teachers made changes in instructional practice during the 2013-2014 and 2014-2015 and the primary impetuses for these changes. Once designed, the survey was field tested with a convenience sample of 40 recruited teachers. Fowler (2014) noted that field testing determines the usability and validity of a survey. When surveys accurately assess the premise of a study with little margin for error, they are deemed valid. Face validity occurred when recruited teachers were able to ascertain that the survey identified classroom instructional practices and measured impetuses for possible changes.

The field test validated that the survey would accurately filter participants, guaranteeing only English, math, social studies, and science teachers of Grades 9-12 could complete the survey. Survey flow, skip logic, and the forced-choice aspects of the survey were analyzed to ensure successful implementation of the survey design. Upon completion of the field-testing period, each participant discussed his or her findings regarding the flow of the survey, the content, and clarity with the researcher.

**Data Analysis and Findings**

As a sequential, mixed-methods, two-phase study, the qualitative phase informed the development of the quantitative phase two. Research questions one and two were addressed in the qualitative phase one of the study, and research questions 3 and 4 were addressed in the quantitative phase of the study.
Research question 1. From the perspective of selected experienced lead evaluators within Central New York school districts who have 2 years of experience under the new APPR process, what are the critical components of classroom instructional practice contained in the state-approved evaluation rubrics? Five lead evaluators with at least 2 years of experience with the APPR process addressed research question 1 during the first focus group. The discussion began with the researcher asking the five lead evaluators to discuss critical components of classroom instructional practice. Each participant joined in the discussion that lasted 75 minutes. The researcher offered follow-up questions and, at times, asked for clarification during the open-ended discussion format.

Coding of the focus-group transcription revealed two central categories for classroom instructional practice: lesson planning and classroom atmosphere. Figures 4.1 and 4.2 illustrate these main categories and their corresponding sub-categories. Sub-categories were identified during the focus group’s discussion.

Figure 4.1 illustrates the components of classroom instructional practice that are associated with planning and delivery of a lesson. Lesson planning encompasses practices that address planning, delivery, assessment, and reflection of instruction.

Five major sub-categories of lesson planning included direct instruction, data-driven instruction, active learning, assessing student learning, and learning goals. Each of these sub-categories interplays with each other and rarely exists independently of one another.

Direct instruction was the part of the lesson where the teacher is presenting information or weaving together components of the lesson. “Looking at instructional
sequencing, such as how much time is direct instruction versus kids being involved in activities,” was reported by a lead evaluator as critical when observing a lesson. Another lead evaluator added, “good teachers know when to stop direct instruction and transition into guided practice, get into a discussion, or some formative assessment that goes along with that.” Direct instruction is the delivery of the lesson’s content as a venue for corollary activities that enrich the lesson.

*Data-driven instruction* was a second sub-category of lesson planning. This instructional practice finds teachers designing, implementing, and revising lessons based upon formative and summative data provided by students. Understanding the cognitive level that each student is performing at and then devising lessons that meet and challenge those levels are at the heart of data-driven instruction. Teachers use data to determine students’ reading levels, and they then differentiate instruction to meet the reading levels at the students’ cognitive levels. Another lead evaluator added, “they have to measure what they are learning every single solitary day with formative assessment in order to know where to take the lesson the next class.”

*Active learning* was a third sub-category of the classroom instructional practice of lesson planning. Active learning is the antithesis of a lecture hall. The teacher in this strategy facilitates learning rather than teaching a lesson. Active learning is best represented by this statement from a lead evaluator, “the children need an opportunity to engage with the material, to make it their own, to personalize it, to question it, and to ask questions about it; they need to get answers, and all of this is done in an environment where the students are actively engaged.” The sage on the stage is replaced with students determining the pacing and paths of their learning. “The opposite of active learning is
Critical Components of Lesson Planning

**Active Learning**
- Engagement of students
- Anticipatory set
- Questioning
- College & career readiness skills
- Technology

**Data-Driven Instruction**
- Level of cognition
- Differentiated learning
- Revision of lesson plans

**Direct Instruction**

**Assessing Student Learning**
- Feedback
- Formative assessments
- Summative assessments
- Closure

**Learning Goals**
- Agendas

*Figure 4.1.* Thematic coding identification of lesson planning as identified by five lead evaluators on July 27 2015.
passive learning, and I think if you go in and watch a teacher who can lecture for 40 minutes, the kids are not going to get that,” stated a lead evaluator while discussing attributes of active learning.

Active learning includes activities that elicit student participation, such as anticipatory set exercises and questioning techniques. Lead evaluators stated active learning includes “students automatically accessing materials to begin the lesson when walking into the classroom,” and “not only raising their hands to answer questions, but also posing their own.”

Assessing student learning was a fourth sub-category of the classroom instructional practice of lesson planning. Assessing student learning is the precursor to data-driven instruction, as this is where the teacher evaluates what the students learned versus what was taught. Assessing student learning includes four main classroom instructional practices. Practices identified as elements of assessing student learning include feedback, formative assessments, summative assessments, and closure.

Assessing student learning is ongoing, and as one lead evaluator stated, “If we wait until the end of the unit to find out if students learned what we expected them to learn, we are kind of doing it post-mortem.” For example, formative assessment noted by one participant “Gives you feedback if you structure the activity correctly.” Another lead evaluator followed up the feedback comment by adding,

one of the things that I think with the ticket out the door is that it gives you feedback if you structure the question right. Whether they met the objectives of the lesson or the goals of the lesson because you are going to get that instant feedback.
Critical with this component is to look at each “individual student along with classroom performance, and performance between sections of classes,” noted a lead evaluator. “Student-centered closure at the end of a lesson or unit allows for direct and timely feedback to the teacher,” according to another lead evaluator. This feedback allows changes to occur immediately so that time is not lost and teachable moments are enhanced.

*Learning goals* was the fifth and last sub-category of the classroom instructional practice of lesson planning. Learning goals provides the roadmap for students through each content topic. This strategy keeps the teacher on task and provides verbal and visual clues for students. Learning goals encompass, as one lead evaluator noted, “intended outcomes, measurable outcomes and activities, learning strategies, and instructional strategies.” The lead evaluator furthered stated, “conveying this to students occurs by posting daily agendas, providing unit maps at the start of each unit, or through posting on the teacher’s webpage.” Simply stated by one lead evaluator, “it provides students with the map for where they are going, a brief overview of how they will get there, and reference for how they will know they have been successful.”

Classroom atmosphere/experience, as represented in Figure 4.2, encompasses the ability of the teacher to develop a professional relationship with the student so that the student is confident that he or she can learn from that teacher. Stated by a lead evaluator, “...education at its barest, at its most fundamental level, is a teacher-to-student relationship.” Each lead evaluator commented on the power of this relationship in terms of engaging students in the learning process.
To assist in assessing this component during the evaluation process, a lead evaluator specifically asks teachers, “Do your kids believe that they can learn from you?” If the teacher answers yes, they are asked to support their answers with evidence. It was often noted that this question, more than all others, causes the greatest pause, and then in-depth conversation occurs about the students and the learning process.

**Critical Components of the Classroom Atmosphere for Instruction**

![Diagram of Classroom Atmosphere](image)

*Figure 4.2.* Thematic coding identification of classroom atmosphere and experience as identified by five lead evaluators on July 27, 2015.

Three major sub-categories of classroom atmosphere included: classroom management, student to teacher credibility, and locus of control within the classroom. All three sub-categories interplay, resulting in the overall tone for the classroom atmosphere.

*Classroom management* was a sub-category of classroom atmosphere. Denoting the importance of classroom management, a lead evaluator stated, “If classroom control doesn’t exist, you can take all of the things on APPR and throw it out the window, because there is going to be no learning taking place.” When asked to explain what that looks like, the participant stated, “You know that the teacher there has expectations, the
teacher has high expectations of student learning, there is a sense of mutual respect between the teacher and student, and that has to be established up front.”

*Student-to-teacher credibility* was a second sub-category of classroom atmosphere. Within this sub-category is the critical component of relationships. A lead evaluator stated, “The first thing I look for is student credibility when I walk into a classroom, and for me credibility is the student believing he can learn from that teacher.” When asked to clarify the meaning of credibility, the lead evaluator stated, “that goes to the relationship, the professional relationship that the teacher establishes and then consistently engages with the kid. Education, at is barest, at its most fundamental level, is a teacher-to-student relationship.”

*Locus of control* was defined by a lead evaluator as “who is in charge of the learning.” Lead evaluators discussed the locus of control as being a fluid component, where the students are able to influence and direct the pace and depth of a lesson. Additionally, a lead evaluator noted that in order to score in the high end on the APPR rubric, the teacher needs to “demonstrate that the students are controlling instruction.”

All five lead evaluators provided insight into the critical classroom instructional practices of effective teachers. They identified 26 critical components of classroom instructional practice.

**Research question 2.** From the perspective of selected teachers, in Grades 9-12 within Central New York school districts, who have 2 years of experience under the new APPR process, what are the primary impetuses for teachers to change classroom instructional practice? Six teachers with at least 2 years of experience with the APPR process addressed research question 2 during the second focus group. The discussion
began with the researcher asking the six teachers to discuss impetuses that caused them to change classroom instructional practice over the past 2 years. Each teacher participated in the discussion that lasted 65 minutes. The researcher offered follow-up questions and at times asked for clarification during the open-ended discussion format.

Coding of the focus group revealed three central categories for impetuses to change classroom instructional practice: feedback, professional development/continuous learning, and planning. Figures 4.3 illustrates these main categories and their corresponding subcategories.

Feedback is an ongoing formal and informal process that provides specific data regarding the teacher’s instructional practice. Feedback is provided in a variety of formats for the classroom teacher as is illustrated in Figure 4.3.

*Anonymous critiques/surveys* was a sub-category of feedback. Some teachers provide surveys to students during the year to assess instructional strategies or the learning environment. One teacher stated that surveys allow an opportunity to “communicate to their students that their ideas are important, and I value their input.” The teachers discussed a desire to create a culture in the classroom where students were willing to advocate for their learning. Students have taken the opportunity on surveys to state, “your notes are really sloppy,” to which a teacher acknowledged the need to provide typed notes.
Categorical Impetuses for Changing Classroom Instructional Practice

<table>
<thead>
<tr>
<th>Feedback</th>
<th>Professional Development / Continuous Learning</th>
<th>Planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Anonymous critiques</td>
<td>• Peer collaboration</td>
<td>• Curriculum changes</td>
</tr>
<tr>
<td>+/- or surveys</td>
<td>• outside of the district</td>
<td>• Common Core curriculum changes from the</td>
</tr>
<tr>
<td>• Formal from administration</td>
<td>• within the school district</td>
<td>State Education Department</td>
</tr>
<tr>
<td>• Informal from</td>
<td>• with student-teachers</td>
<td>• Personal experience</td>
</tr>
<tr>
<td>administration</td>
<td>• Professional development offered by the school</td>
<td>as a parent</td>
</tr>
<tr>
<td>• APPR process</td>
<td>district (district initiatives or in-services)</td>
<td></td>
</tr>
<tr>
<td>• From parents</td>
<td>• Conferences outside of the district with a</td>
<td></td>
</tr>
<tr>
<td>• From students</td>
<td>K-12 focus</td>
<td></td>
</tr>
<tr>
<td>• Student performance</td>
<td>• Content/grade level specific conferences</td>
<td></td>
</tr>
<tr>
<td>• summative assessments</td>
<td>• Professional journals or articles</td>
<td></td>
</tr>
<tr>
<td>• formative assessments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Re-evaluate and re-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>assess during classroom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>instruction</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.3.** Data identified from coding teacher focus group discussion on what has caused them to change classroom instructional practice over the past 2 years.

*Formal from administration* was a sub-category of feedback. Formal feedback was stated by a teacher as causing change because the “administrator told me to focus more on student-centered closure and less on my review of a lesson.” This impetus occurs when an administrator presents a specific strategy or practice that is part of a district initiative.

*Informal from administration* was a sub-category of feedback. A couple of teachers stated formal and informal feedback from administrators caused them to change practice. “So when an administrator comes in and watches my anticipatory set and then later in the day discusses a different way a similar teacher opened their class, I may be
inclined to try the new idea” was offered by a teacher as an example of informal feedback from an administrator.

**APPR process** was a sub-category of feedback. Of note, only one teacher mentioned the teacher evaluation process, APPR, as an impetus for change, “…yeah, sometimes feedback from the APPR process helps, but I think that some of the things that have happened that aren’t strictly part of the APPR process have been more valuable.”

**Parent feedback** was a sub-category of feedback. Feedback may not always contain a positive message, but does provide a context for change in the classroom. One teacher noted, “It was a little unnerving to have a parent tell you that their child struggled, you know, going into a calculus course because of the lack of the use of calculators in college.” Resulting from this informal comment, the teacher changed instruction in the pre-calculus class so that it focused more on the theory and the math and less on the incorporation of calculators.

**Student feedback** was a sub-category of feedback. Students are encouraged to offer suggestions and ideas to the teacher to help their learning or the learning environment. One teacher stated,

I try to create a culture where they can, whether it is in the middle of the class, they can say Mr. Smith, this isn’t working for me, or they can stop and see me after class, or send me an email. I try and create a culture where students feel that they can provide feedback about the classroom environment, such as what classroom activities are working, or you need to be more clear with your notes. Regarding parent and student feedback, student feedback causes more change than parental.
“There is seldom classroom-wide changes that are made as a result of parent feedback” stated one teacher. The same teacher added, “The biggest changes have been caused primarily from feedback from my students, and also second of all, from my administrators.”

Student performance was a sub-category of feedback. This sub-category included both summative and formative assessments. Formative assessments are ongoing throughout a unit of study, whereas summative assessments include the final assessments for a unit of study or for the course. One teacher stated that when students score poorly on a quiz,

I then know that is definitely a topic that just did not work well, so I go back to that day’s lesson and think back about the lesson and say, ‘ok, did I not have enough opportunity for them to practice this? Was my homework assignment not good enough? Was the test not aligned properly to with what was taught in the lesson?’ And those are the things I consider when making changes.

Re-evaluate and re-assess was a sub-category of feedback. Student enjoyment with a particular unit, or lack thereof, caused some teachers to change practice. Being assigned a new prep caused one teacher to look at “why we do things this way” regarding activities to support a unit. This teacher furthered, doing something brand new in the classroom, and to re-introduce, re-organize my classroom in a completely different way and restructure it, felt fresh and invigorating. I think my enthusiasm kind of rubbed off on the class, and they were enthusiastic about it.

Five sub-categories of Professional Development/Continuous Learning included peer collaboration, professional development offered by the school district, conferences
outside the school district with a K-12 focus, content/grade-level-specific conferences, and professional journals or articles. Teachers are part of a career path that requires them to be lifelong learners, and as such, they have a variety of opportunities for professional development.

Peer collaboration was a sub-category of the professional development/continuous learning. Peer collaboration occurs between teachers within a school district, between different districts, and with student teachers. Working with a student teacher was an impetus that one teacher had experienced. The experience caused the teacher to reevaluate and reassess classroom instructional practices. Specifically, the teacher stated, “having the student teacher come in and ask why do you do things this way?” was an impetus for changing classroom instructional practice. Having a co-teacher in the classroom also led teachers to change practice. “A lot of suggestions came from my co-teacher as a result of conferences attended. She provided feedback to me regarding the lessons, and we then worked together to develop more effective lesson plans.” Another teacher added to the concept of co-teaching, having another teacher in my room to share space with teaching and just the level of student, it was a whole eye opener, they are not doing their homework, it’s just that they don’t understand. Gosh, I had to change.

Professional development offered by the school district was a sub-category of professional development/continuous learning. The theory behind teacher evaluation is that it drives changes in classroom instructional practice and sets the direction of professional development. It was not surprising that impetuses for change were often associated with professional development. District initiatives also lead teachers to
change instructional practice. “…this year I had a co-teaching class, and that facilitated a lot of change in my classroom,” stated one teacher.

*Conferences outside the school district with a K-12 focus* was a sub-category of professional development/continuous learning. This impetus was mentioned in relation to training for co-teaching and received little discussion.

*Content/grade-level-specific conferences* was a sub-category of professional development/continuous learning. This impetus received more discussion than conferences outside of the district. A teacher stated, “another thing that I feel has been really helpful for me is that we have been attending a lot of conferences with someone who is very knowledgeable about what is going on with the changes to the math curriculum.” Attending training specific to the content taught led to changes to this teacher’s classroom practice.

*Professional journals or articles* was a sub-category of the professional development/continuous learning. Discussion on this impetus led one teacher to state, “reading professional journals is an important part of ongoing improvement, it’s just reading, you know, whether it is from professional journals, or just something in the popular media that you read about education. Ongoing is critical.”

Two sub-categories of planning included *curriculum changes from the State Education Department* and *personal experience as a parent*. This category addressed two critical aspects of the overall educational process, the first being curriculum changes from the New York State Department of Education and the second including personal awareness of how children learn.
Curriculum changes was a sub-category of planning. New York State Education Department adopted the Common Core Learning Standards and their implementation has been the focal point for school districts. It has impacted teachers as noted, “in math we’ve had a lot of curriculum changes the past 2 years, so I think, initially, that’s the biggest thing that has caused us to change things in the classroom.” Another teacher echoed, “the state Common Core curriculum has drastically affected the changes in my classroom as of recently.”

Personal experience as a parent was a sub-category of planning. Being a parent was an impetus that prompted discussion by every teacher. A teacher began the discussion by stating, “Some of the things that I’ve changed in my classroom is from watching some of the struggles that go on at home.” From this statement began a discussion on how children learn and what teachers need to do in order to address everyone fairly and equally. For example, “I am much more willing to give them an extension on their homework than I probably was 10 years ago,” and “I just have to find different ways to present the material” so everyone can learn. A teacher who is not a parent furthered the discussion by noting the corollary to being a coach rather than a parent. Through this lens, the teacher viewed how students learn and internalize information differently. Being a coach was noted as being the equivalent of being a parent in regard to causing change within the classroom.

Research question 3. Given the critical components of classroom practice, as identified by question 1, from the perspective of selected English, math, science, and social studies teachers in Grades 9-12 within Central New York school districts, in which component(s) did they make the most changes in instructional practice over the past 24
months? The data collected from both focus groups in phase one led to the creation of a 10-question survey using Qualtrics online software for phase two of the study. The first six questions were demographic to filter the participation down to English, math, science, and social studies teachers of Grades 9-12. Figure 4.4 illustrates teacher participation by subject area in the survey. Skip logic was incorporated to ensure the participation of only these specific teachers.

**Subject Areas Taught by Participants**

![Subject Areas Taught by Participants](image)

*Figure 4.4.* The study included an even distribution of participation from the four core subject areas.

The survey was open during a 14-day period from September 14, 2015 to September 28, 2015. Of the 27 public school districts invited to participate in the study, 18 accepted, for a participation rate of 66%. An email containing an electronic link to the survey, via Qualtrics.com, was sent to all participating districts. At the close of research solicitation, 292 teachers from the 18 public school districts accessed the survey. Initial questions narrowed participation in the survey to English, math, science, and social studies teachers of Grades 9-12, resulting in 114 participants. Of the 114 participants, seven reported that they did not have 2 years of evaluation experience as a classroom teacher. Participation furthered narrowed as 21 teachers reported that they did not make
any changes in classroom instructional practices during the 2013-2014 or 2014-2015 school year, thus narrowing participation in the survey to 83 participants.

Lead evaluators identified two critical components of classroom instruction: *classroom lesson planning* and *classroom atmosphere design*. Classroom lesson planning was defined as including the actual lesson plan formatting, as well as any or all of the design and implementation of the lesson plan, direct instruction, language acquisition, teaching vocabulary, data-driven instruction, identifying and focusing on student levels of cognition, differentiated learning, lesson plan revision, activities and strategies to improve student engagement, anticipatory sets, incorporation of technology, formative and summative assessments, feedback to students, learning goals, and class agendas.

The second critical component of classroom instructional practice identified was classroom atmosphere. Classroom atmosphere was defined as any one or all of the manner in which the classroom management practices are designed and implemented, the rapport developed between student and teacher, the relationship built between the teachers and their classes, the relationship built amongst students, and the fluid movement to teacher as facilitator where students control the pacing and path of their learning.

Lesson planning was the classroom instructional component where teachers made the most changes during the 2013-2014 and 2014-2015 school years. Table 4.1 illustrates the responses reported by the participants regarding making changes in classroom instructional practice.
Table 4.1

Changes Reported in Classroom Instructional Practices

<table>
<thead>
<tr>
<th>Instructional Practice</th>
<th>Answer</th>
<th>Response</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson Planning</td>
<td>Yes</td>
<td>83</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Classroom Atmosphere</td>
<td>Yes</td>
<td>54</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>26</td>
<td>33</td>
</tr>
</tbody>
</table>

Note. Responses addressed whether changes were made in classroom instructional practices during the 2013-2014 or 2014-2015 school years.

Research question 4. Given the changes identified by the same participants in question 3, what was the primary impetus for changes made in their classroom instructional practice? Figure 4.5 provides a visual representation of the degree to which the participants changed their classroom instructional practices associated with lesson planning. The top bar identifies no change, the middle bar identifies moderate change, and the bottom bar represents significant change. Table 4.2 follows with a statistical representation including the mean (μ), standard deviation (SD), and variance (s²).

Teachers rarely reported making significant changes in lesson planning based upon the list of impetuses provided. The mean averages 1 point for no change, 2 points for moderate change, and 3 points for significant change. Thus, the mean provides a holistic view of changes based upon the impetus. Table 4.2 provides data identifying six primary impetuses for changing classroom instructional practice as represented by the mean. The three categorical ratings for the table are in regard to the changes reported: none, moderate, and significant.
Figure 4.5. Degree to which participants changed classroom instructional practice based upon the list of identified impetuses. None is depicted in the left column, moderate is the middle column, and significant is in the right column.

Table 4.2 also denotes the placement of the APPR process as an impetus for causing change in teachers’ classroom instructional practice focusing on lesson planning. Just over 50% of teachers acknowledged making changes to lesson planning due to the formal APPR process in place within their school district ($\mu = 1.66$). The top line
indicates no change, the middle line indicates moderate change, and the bottom line indicates significant change.

Table 4.2

*Top Six Impetuses For Changing Lesson Planning*

<table>
<thead>
<tr>
<th>Impetus</th>
<th>None</th>
<th>Moderate</th>
<th>Significant</th>
<th>μ</th>
<th>SD</th>
<th>s²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborate with peers within your own school district</td>
<td>13</td>
<td>45</td>
<td>19</td>
<td>2.08</td>
<td>.64</td>
<td>.41</td>
</tr>
<tr>
<td>Student performance during class time/instruction</td>
<td>15</td>
<td>49</td>
<td>13</td>
<td>1.97</td>
<td>.61</td>
<td>.37</td>
</tr>
<tr>
<td>Students performance on formative assessments during the school year</td>
<td>16</td>
<td>49</td>
<td>12</td>
<td>1.95</td>
<td>.60</td>
<td>.37</td>
</tr>
<tr>
<td>Curriculum changes resulting from Common Core implementation</td>
<td>24</td>
<td>33</td>
<td>20</td>
<td>1.95</td>
<td>.76</td>
<td>.58</td>
</tr>
<tr>
<td>Student performance on summative assessments during</td>
<td>18</td>
<td>46</td>
<td>13</td>
<td>1.94</td>
<td>.64</td>
<td>.40</td>
</tr>
<tr>
<td>Ongoing school district initiatives</td>
<td>13</td>
<td>56</td>
<td>8</td>
<td>1.94</td>
<td>.52</td>
<td>.27</td>
</tr>
<tr>
<td>Directly resulting from the APPR process at your school</td>
<td>35</td>
<td>33</td>
<td>9</td>
<td>1.66</td>
<td>.68</td>
<td>.46</td>
</tr>
</tbody>
</table>

*Note.* Mean was determined by scoring none with one point, moderate with two, and significant changes with 3 points. M = mean; SD = standard deviation; and s² = variance.

Variation and standard deviation are both measures of the spread of values in a data set. Creswell (2013) defined variance as the measure of dispersion of the values from the mean. It accounts for the variation from the average data point and is known as the squared value of the mean. Variance is represented with the symbol s². The squared root of variance results in the standard deviation (Creswell, 2013), a metric that identifies outliers in the data set. The standard deviation statistic symbol is represented as SD.

Collaboration with peers within one’s own school district and student performance during class time were identified as primary impetuses for causing teachers
to change practice. Collaboration with peers scored the highest mean ($\mu = 2.08$), and student performance during class time instruction scored the second highest mean ($\mu = 21.97$), with a slightly lower standard deviation (SD = .61). Little variance was reported among the scores for both impetuses, collaboration with peers ($s^2 = .41$) and student performance during class time instruction ($s^2 = .37$). Of the participants, 81% reported making changes in lesson planning based on both impetuses.

Student performance on formative assessments, curriculum changes resulting from Common Core implementation, student performance on formative assessments during the year, and student performance on summative assessments during the year accounted for the remainder of the top six impetuses. Each of these impetuses found the mean within .03 of one another. Standard deviation and variance were similar for all impetuses except for curriculum changes resulting from Common Core implementation, which was higher. The variance ($s^2 = .58$) is the highest for all impetuses, as 31% of the respondents did not make changes based upon Common Core implementation compared to 68% who did make changes.

Table 4.3 presents the ranking of the remaining 14 impetuses based upon the mean. Similarly, the data reveals low mean ($\mu$) scores with high variation ($s^2$) scores indicate a fairly even distribution of teachers identifying the impetus as not effecting any change in classroom practice.

Collaboration with student teachers found low variance ($s^2 = .12$) and a tight standard deviation (SD = .35), which when correlated against the mean ($\mu = 1.11$), it is clearly indicative that this impetus does not motivate teachers to change practice. Teachers were practically unanimous in this determination as they were with regard to
feedback from parents (SD = .42, s² = .17) and a low mean (μ = 1.22). This impetus was
the only category where teachers did not record any significant changes occurring from
this impetus.

Table 4.3

*Impetuses for Changing Classroom Instructional Practices in Lesson Planning*

<table>
<thead>
<tr>
<th>Impetus</th>
<th>None</th>
<th>Moderate</th>
<th>Significant</th>
<th>M</th>
<th>SD</th>
<th>s²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum changes within the school district</td>
<td>28</td>
<td>33</td>
<td>16</td>
<td>1.84</td>
<td>.74</td>
<td>55</td>
</tr>
<tr>
<td>Professional development offered by your own school district</td>
<td>24</td>
<td>42</td>
<td>11</td>
<td>1.83</td>
<td>.66</td>
<td>.43</td>
</tr>
<tr>
<td>Informal feedback from students</td>
<td>21</td>
<td>50</td>
<td>6</td>
<td>1.81</td>
<td>.56</td>
<td>.32</td>
</tr>
<tr>
<td>Formal feedback from Administrators (outside APPR)</td>
<td>32</td>
<td>33</td>
<td>12</td>
<td>1.74</td>
<td>.71</td>
<td>.51</td>
</tr>
<tr>
<td>Informal feedback from administrators (outside APPR)</td>
<td>32</td>
<td>37</td>
<td>8</td>
<td>1.69</td>
<td>.65</td>
<td>.43</td>
</tr>
<tr>
<td>Attendance at content/grade level specific conferences</td>
<td>32</td>
<td>39</td>
<td>6</td>
<td>1.66</td>
<td>.62</td>
<td>.38</td>
</tr>
<tr>
<td>Collaborate with peers outside of your own school district</td>
<td>41</td>
<td>30</td>
<td>6</td>
<td>1.55</td>
<td>.64</td>
<td>.41</td>
</tr>
<tr>
<td>Attendance at conferences geared towards K-12 education</td>
<td>38</td>
<td>36</td>
<td>3</td>
<td>1.55</td>
<td>.57</td>
<td>.33</td>
</tr>
<tr>
<td>Articles in professional journals or books written about instructional practices</td>
<td>46</td>
<td>27</td>
<td>4</td>
<td>1.45</td>
<td>.60</td>
<td>.36</td>
</tr>
<tr>
<td>Anonymous surveys/critiques from students</td>
<td>48</td>
<td>25</td>
<td>4</td>
<td>1.43</td>
<td>.59</td>
<td>.35</td>
</tr>
<tr>
<td>Personal experience as a parent</td>
<td>54</td>
<td>16</td>
<td>7</td>
<td>1.39</td>
<td>.65</td>
<td>.43</td>
</tr>
<tr>
<td>Feedback from parents</td>
<td>60</td>
<td>17</td>
<td>0</td>
<td>1.22</td>
<td>.42</td>
<td>.17</td>
</tr>
<tr>
<td>Collaboration with student teachers</td>
<td>70</td>
<td>6</td>
<td>1</td>
<td>1.10</td>
<td>.35</td>
<td>.12</td>
</tr>
</tbody>
</table>

*Note.* Mean was determined by scoring none with one point, moderate with two, and
significant changes with 3 points. M = mean; SD = standard deviation; and s² = variance.
Figure 4.6 provides a visual representation regarding changes in the classroom atmosphere. Of the participants, 54 reported making changes in the instructional practice of classroom atmosphere. Table 4.4 follows with a statistical representation including the mean (μ), standard deviation (SD), and variance (s²). The top line indicates no change, the middle line indicates moderate change, and the bottom line indicates significant change.

![Reported Impetuses for Changing Classroom Atmosphere](image)

**Figure 4.6.** Degree to which participants made changes to the classroom instructional practice, the classroom atmosphere, based upon the list of identified impetuses. None is depicted in the left column, moderate is the middle column, and significant is in the right column.
Significant changes were recorded for several impetuses, with curriculum changes as a result of Common Core implementation and student performance during class time and instruction earning the highest scores. With the highest mean ($\mu = 2.07$), and lower variance ($s^2 = .33$), it is clear that student performance during class time instruction caused teachers to change practice. Feedback from parents and collaboration with student teachers did not result in any significant changes as recorded by the participants. Working with student teachers found no significant variance ($s^2 = .07$) with an equally insignificant standard deviation (SD = .26). Table 4.4 provides data identifying five primary impetuses for changing classroom atmosphere. The three categorical ratings for the table are in regard to the changes reported: none, moderate, and significant.

Table 4.4 also denotes the placement of the APPR process as an impetus for causing change in teachers’ classroom instructional practice focusing on lesson planning. Just over 50% of teachers acknowledge making changes to lesson planning due to the formal APPR process in place within their school district ($\mu = 1.66$).

Fourth place was a tie, as student performance on formative assessments during the school year, and curriculum changes within the school district, received the exact same scores. When combining moderate and significant change data, 70% of teachers made changes based upon eight impetuses. Table 4.4 also denotes the placement of the APPR process as an impetus for causing change in teacher’s classroom instructional practice focusing on lesson planning. Just under 50% of teachers acknowledged making changes to lesson planning due to the formal APPR process in place within their school district ($\mu = 1.50$).
Table 4.4

*Top Five Impetuses For Changing Classroom Atmosphere*

<table>
<thead>
<tr>
<th>Impetus</th>
<th>None</th>
<th>Moderate</th>
<th>Significant</th>
<th>μ</th>
<th>SD</th>
<th>s²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student performance during class time/instruction</td>
<td>7</td>
<td>36</td>
<td>11</td>
<td>2.07</td>
<td>.58</td>
<td>.33</td>
</tr>
<tr>
<td>Collaborate with peers within your own school district</td>
<td>12</td>
<td>33</td>
<td>9</td>
<td>1.94</td>
<td>.63</td>
<td>.31</td>
</tr>
<tr>
<td>Informal feedback from students</td>
<td>13</td>
<td>37</td>
<td>4</td>
<td>1.83</td>
<td>.54</td>
<td>.21</td>
</tr>
<tr>
<td>Student performance on formative assessments during school year</td>
<td>17</td>
<td>30</td>
<td>7</td>
<td>1.81</td>
<td>.65</td>
<td>.42</td>
</tr>
<tr>
<td>Curriculum changes within the school district</td>
<td>17</td>
<td>30</td>
<td>7</td>
<td>1.81</td>
<td>.65</td>
<td>.42</td>
</tr>
<tr>
<td>Ongoing school district initiatives</td>
<td>16</td>
<td>35</td>
<td>3</td>
<td>1.76</td>
<td>.30</td>
<td>.55</td>
</tr>
<tr>
<td>Directly resulting from the APPR process at your school</td>
<td>30</td>
<td>21</td>
<td>3</td>
<td>1.50</td>
<td>.61</td>
<td>.37</td>
</tr>
</tbody>
</table>

*Note.* Mean was determined by scoring none with one point, moderate with two, and significant changes with 3 points. M = mean; SD = standard deviation; and s² = variance.

Table 4.5 presents the ranking of the remaining impetuses. Similarly, the data revealed low mean (μ) scores with high variation (s²) scores indicating a relatively even distribution of teachers identifying the impetus as causing change. More than half of the respondents made changes to classroom atmosphere instructional practices based upon the top five impetuses on this chart. These impetuses significantly drop off as a cause for teachers to change practice midway through the table with a mean of μ = 1.39 and less. Collaboration with student teachers was reported as not an impetus for change as denoted with the μ = 1.07, and significantly no variance (σ = .07).
Table 4.5

*Impetuses for Changing Classroom Instructional Practices in Classroom Atmosphere*

<table>
<thead>
<tr>
<th>Impetus</th>
<th>None</th>
<th>Moderate</th>
<th>Significant</th>
<th>M</th>
<th>SD</th>
<th>s²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum changes as a result of Common Core implementation</td>
<td>19</td>
<td>23</td>
<td>12</td>
<td>1.87</td>
<td>.75</td>
<td>.57</td>
</tr>
<tr>
<td>Student performance on summative assessments during the school year</td>
<td>18</td>
<td>27</td>
<td>9</td>
<td>1.83</td>
<td>.69</td>
<td>.48</td>
</tr>
<tr>
<td>Informal feedback from administrators (outside APPR)</td>
<td>21</td>
<td>27</td>
<td>6</td>
<td>1.72</td>
<td>.66</td>
<td>.43</td>
</tr>
<tr>
<td>Professional development offered by your own district</td>
<td>18</td>
<td>33</td>
<td>3</td>
<td>1.72</td>
<td>.56</td>
<td>.32</td>
</tr>
<tr>
<td>Formal feedback from administrators (outside APPR)</td>
<td>23</td>
<td>24</td>
<td>7</td>
<td>1.70</td>
<td>.69</td>
<td>.48</td>
</tr>
<tr>
<td>Anonymous surveys and critiques from students</td>
<td>27</td>
<td>25</td>
<td>2</td>
<td>1.54</td>
<td>.57</td>
<td>.33</td>
</tr>
<tr>
<td>Attendance at conferences geared towards K-12 education</td>
<td>29</td>
<td>24</td>
<td>1</td>
<td>1.48</td>
<td>.54</td>
<td>.29</td>
</tr>
<tr>
<td>Collaborate with peers outside of your own school district</td>
<td>31</td>
<td>21</td>
<td>2</td>
<td>1.46</td>
<td>.57</td>
<td>.33</td>
</tr>
<tr>
<td>Personal experiences as a parent</td>
<td>37</td>
<td>13</td>
<td>4</td>
<td>1.39</td>
<td>.63</td>
<td>.39</td>
</tr>
<tr>
<td>Articles in professional journals or books written about</td>
<td>37</td>
<td>14</td>
<td>3</td>
<td>1.39</td>
<td>.63</td>
<td>.39</td>
</tr>
<tr>
<td>Feedback from parents</td>
<td>36</td>
<td>18</td>
<td>0</td>
<td>1.33</td>
<td>.48</td>
<td>.23</td>
</tr>
<tr>
<td>Collaboration with student teachers</td>
<td>50</td>
<td>4</td>
<td>0</td>
<td>1.07</td>
<td>.26</td>
<td>.07</td>
</tr>
</tbody>
</table>

*Note.* Mean was determined by scoring none with one point, moderate with two, and significant changes with 3 points. Mean was determined by scoring none with one point, moderate with 2, and significant changes with 3 points. M = mean; SD = standard deviation; and $s^2$ = variance.

The primary impetuses causing teachers to change classroom instructional practice were similar for both lesson planning and classroom atmosphere. Collaboration with peers is the number one impetus for change in lesson planning, closely followed by student performance during class-time instruction. The inverse is true for classroom atmosphere where student performance during class-time instruction lead to collaboration with peers. Classroom instructional practices of lesson planning and classroom
atmosphere included four identical impetuses for change in the top six reported by teachers.

**Summary of the Results**

From the perspective of five selected lead evaluators from Central New York school districts, 20 critical components of classroom instructional practice were identified. The 20 instructional practices emerged into two distinct categories: lesson planning and classroom atmosphere. From the perspective of six selected teachers from Central New York school districts, 20 impetuses causing teachers to change classroom instructional practice were identified. Lesson planning was identified as the critical component of instructional practice where changes occurred, as reported by classroom teachers.

Chapter 5 delves into the implications of the findings from the study and includes limitations and recommendations for further research. Also included in Chapter 5 are suggestions for policy development based upon the data revealed in the study and recommendations for change in practice.
Chapter 5: Discussion

Introduction

Teacher evaluations have received scholarly attention for decades in an effort to improve student learning. Kersten and Israel (2005) stated that the call for school improvement began with the publishing of *A Nation at Risk* in 1983. The premise from this study offers that holding teachers accountable for student learning will improve student performance. High standards and accountability for teachers continued with the passing of the Elementary and Secondary Education Act in 1965. Congressional reauthorization of this act continues to hold teachers accountable for student learning.

New York State was awarded $696,646,000 from the federal Race to the Top competitive grant program in August 2010 (NYSED, 2010). In compliance, New York Stat passed Education Law §3201-c, which changed the conduct of the APPR process for teachers (EngageNY, 2013). Revised in June 2015, Education Law §3201-d incorporates student performance on assessments and teacher evaluation, equally, when calculating the final rating for teachers.

The Annual Professional Performance Review is a rigorous teacher evaluation system that holds teachers accountable for student learning. Teacher evaluation is one of the most important and complex tools in education (Bereens, 2000), and it is imperative that it does more than just identify strong and weak teachers (Papay, 2012). To be effective, teacher evaluation tools must assist teachers in assessing their performance and provide a road map for developing pedagogical skills that improve achievement for each
student. Recognizing the need for formative tools to assist teachers in continuous improvement, the study explored the extent to which APPR promotes teachers to change classroom instructional practice.

Performance feedback theory served as the framework for the study. In this chapter the implications of the findings are discussed and interpreted as it relates to the impetuses promoting change by teachers in classroom instructional practice. Secondly, the connection between APPR and performance feedback theory is presented. Thirdly, limitations to the study are discussed, followed by recommendations for further research, policy development, and practice. Finally, connections to literature and the theoretical rationale are presented alongside a review of the methodology, results, and conclusions.

Implications of Findings

The study of APPR, as instituted by New York State to evaluate teachers in an effort to improve student performance, has provided meaningful insight into the impetuses for teachers to change instructional practice in the classroom. It also reveals a challenge faced by state educational officials and school district administrators as they implement APPR, as required by law, in an effort to improve student performance. Lastly, evidence from the study stands in contrast to the tenets of Greve’s (2003) performance feedback theory that touts employee evaluation as an impetus for changing practice.

Lead evaluator focus group. Lead evaluators met during a focus group meeting to commence the study. During their discussion on the critical components of classroom instructional practice, they were uniform in their analysis of best practice. When one lead evaluator was speaking, others projected their agreement through nodding their heads and
occasional verbal affirmations. Whereas the lead evaluators spoke in reference to three
different state-approved evaluation rubrics, they clearly delineated critical components of
instructional practice as a whole. Consistency between the available rubrics provides
school districts with a clear description of research-based classroom instructional
practices. The rubrics provide clear indicators regarding what elements are necessary for
effective lessons for classroom teachers.

**Teacher focus group.** Teacher discussion during the focus group was robust and
filled with excitement. It was evident that the teachers were serious about their
profession and enjoyed discussing their practice. One interesting topic discussed during
the focus group meeting, but that did not find significant support in the survey, included
the role of the teacher as a parent or coach.

Teacher discussion in the focus group about their role as a parent or a coach
affirmed the Runhaar et al. (2010) findings concerning a teacher’s willingness to change
and his or her motivation to do so. Specifically, teachers’ noted that their roles as a
parent or a coach enabled them with opportunities to witness student learning styles and
struggles as well as the techniques and practices necessary to help each child grow. In
turn, these techniques and practices transferred from the home and playing field to the
classroom. The classroom teachers realized the need to accept that students require
different scaffolding to succeed, and patience is needed at times on the part of the
teacher. Personal experience with this facet of student learning promoted changes in the
classroom.

**Lesson planning.** The capstone of the study included a survey that captured the
teachers’ impetuses for changing practice in the classroom. Results from the survey
clearly illustrated that the practices associated with lesson planning was where teachers made the most changes in instructional practice. Given the depth of curriculum changes required by implementation of the national Common Core standards, changes in state assessments, and changes to College Board standardized assessments, including Advanced Placement exams and Scholastic Aptitude Tests, it was not surprising for this category of instructional practice to receive the most changes recorded by the teachers.

**Collaboration with peers.** Collaboration with peers within their own school district received the highest marks as impetus for changing the classroom instructional practice of lesson planning. Whereas the five other main impetuses for changing lesson planning were within a tight range of one another, collaboration with peers within their own school district was notably a primary choice by the teachers. This has significant implications for professional development within school districts, as well as regarding the master scheduling process. School districts have an opportunity to maximize this impetus by providing systemic opportunities within the school day and within the school year for teachers to work collaboratively with one another. Common planning time and time reserved during professional development days would allow the teachers to have an opportunity to collaborate at no fiscal cost to the school district.

**Student performance during class time instruction.** This impetus ranked second for promoting the teachers to change instructional practice related to lesson planning. Formal and informal feedback from students that occurs while the teacher is delivering the lesson drove the teachers to make changes. Being able to recognize an issue and knowing what to do to either enhance or change student learning is organic to teachers. Knowledge possessed by the teacher develops with years of service, through
collaboration with peers, and through staff development provided by the district, none of which cost the school district significant money.

**APPR.** New York State is paying close to $700,000,000 per year to implement the APPR system. APPR ranked ninth out of 20 on a list of impetuses that promote change by classroom teachers. Collaborating with peers within their own school district ranked number one as an impetus for causing teachers to change practice in the classroom. Providing time for teachers to collaborate with one another does not require fiscal expenditures by the state nor the school district.

Hazi (2014) stated that there is no link between the teacher evaluation process and student learning. However, Darling-Hammond (2013a), Danielson (2001), and Tucker and Stronge (2005) stated that a strong relationship exists between student learning and the quality of the teacher. Part of the disconnect between teacher evaluations and student learning lies in the complexity of teacher evaluation systems (Callahan & Sadeghi, 2015). Most systems serve a multitude of purposes, resulting in a poor tool for analyzing teacher quality. Motivation to change practice occurs between teachers in a collegial environment and not as part of a complex system designed to sort teacher performance.

**Classroom atmosphere.** The second category identified by lead evaluators as a critical component of classroom instructional practice was the classroom atmosphere. Impetuses for changes within this category included the same top two choices, although reversed. Student performance during class time instruction was identified as the primary impetus, and collaboration with peers within the same school district was second. The implications of these findings is substantial as school districts can equally address both categories of classroom instructional practice by focusing on making time available for
peers to collaborate and be involved in professional development on how to maximize formal and informal formative feedback from students.

**APPR in New York State.** Implementation of APPR in New York State occurred as the Governor David Patterson and State Assembly leaders created an evaluation law in order to qualify for federal grant money available through Race to the Top. Implemented in 2010, the plan has experienced annual revisions to improve its effectiveness and acceptance by teachers and school leaders. Pearce and Porter (1986) acknowledged that attitudes toward the evaluation performance system impacts employees’ commitment toward the process.

Discussion among the teachers regarding impetuses to changing classroom instructional practice easily passed one hour before mentioning APPR. This phenomenon is in stark contrast to Ponticell and Zepeda’s (2004) findings that adults are performance-centered and this drives their desire for continuous improvement. In the study, teachers clearly acknowledged the desire for continuous learning and improvement, but they did not offer the APPR system as providing the impetus for doing so. Taken further, the findings from this study support the theoretical framework provided by Greve’s (2003) performance feedback theory, although in a less formal manner.

According to performance feedback theory, when employees receive feedback regarding their performance, they will choose whether or not to enact changes to improve practice. Feedback may be comparative data within the organization or between similar organizations, summative results of performance, or other metrics associated with the organization. The study reports teachers seeking feedback from one another in a
collaborative spirit as an impetus to change practice. The impetus is a less formal source than those purported by the performance feedback theory.

Constructive feedback is a key component of the newly implemented APPR system in New York State. As indicated by this study, feedback from the APPR process, however, did not drive the teachers to change their practice. Feedback from student performance during class time instruction did promote the teachers to change their practice. Collaboration with peers and feedback from student performance during class time provided information to the teachers. Information was then assessed to determine whether to change classroom instructional practice. With this cyclical paradigm in play, the process mimics performance feedback theory.

**Performance feedback theory.** Greve’s (2003) performance feedback theory posits that employees set aspiration levels, or goals, based on past performances. Job performance changes to meet aspiration levels, data is subsequently collected to determine if the aspiration level was met, and then new aspiration levels are established for continuous improvement. The process is cyclical and continuous using these three main aspects: collect performance measures, create aspiration levels, and change activities if performance is low. One caveat to this cycle is satisficing.

Satisficing occurs when the performance level meets the aspirational goal, thus resulting in no change. It may also occur when the feedback indicates that the performance level is just beneath the aspiration goal level, thus resulting in very little change. The optimal target is to reach the aspiration goal and to not overachieve it.

Teaching is a natural application of performance feedback theory, as the classroom teacher uses data and past performance to set classroom and student
expectations. Ongoing assessments during class time instruction informs the teacher regarding student learning. Changes occur when students are not successful, and the cycle continues throughout the year, as the teacher modifies instruction to meet students’ needs. In fact, the cyclical nature is yearly for students, but for the classroom teacher, the cycle of changing is ongoing throughout their career.

It was interesting that 21 of the 114 participating teachers reported not making any changes to their classroom instructional practices during the 2012-2013 and 2013-2014 school years. Applying performance feedback theory to this statistic would indicate that possibly these 20% of teachers viewed their performance feedback as acceptable, and they did not employ changes. Satisficing is apparent here, as changes would have led to exceeding the aspirational goals, and they were comfortable not exceeding the mark.

**Limitations**

One limitation that may have impacted the study concerns the timing of the delivery of the survey to the teachers. An ideal timing for delivery of the survey would have coincided with the close of the school year when high school teachers received their final composite APPR scores. Instead of receiving the survey in June, English, math, science, and social studies teachers of Grades 9-12 received the survey in late September. In September teachers are busy with the opening of school and numerous commitments associated with getting the school year underway. At this time of year, the frame of reference for teachers is far removed from APPR, and this may have affected teacher participation in the survey.
Recommendations

The study focused on the extent to which the APPR process promoted change in classroom instructional practice by the teachers. Investigating this phenomenon included identification of critical components of classroom instructional practice and identification of impetuses for changing classroom instructional practice. The study found that lesson planning is where most classroom change occurs, and collaboration with peers within their school district is the impetus resulting in change in order to improve student performance. These are valuable tools of knowledge for school districts as they seek to improve student performance.

Recommendation for further research. Whereas the participation in the study focused primarily on English, math, science, and social studies teachers of Grades 9-12, a recommendation for future research is to use a similar lens to study elementary or middle school teachers. Elementary and middle school education is primarily skills-based, in comparison to the content-based high school education, therefore posing an interesting contrast regarding what causes teachers to change practice.

A second recommendation would be to conduct the study with a focus on schools that earned the New York State Rewards School designation, and compare them to schools that earned the Priority Schools designation. The following criteria must be met to be considered a Reward School: (a) Be among the top 20% of schools in the state for English language arts and math performance or be among the top 10% of schools in terms of gains in ELA and math performance; (b) made adequate yearly progress for all groups of students on all measures for which the school is accountable, including the requirement that 95% of all groups participate in the English language arts and
mathematics assessments; and (c) large gaps cannot exist between accountability groups for the district (New York State Education Department [NYSED], 2014c). Priority schools include the lowest performing 5% of Title I schools for the past 3 years (NYSED, 2015c). Studying Rewards Schools would identify the impetuses for changing practice resulting in high student achievement. Conversely, studying Priority Schools would give districts the knowledge on how best to design professional develop for their staff in order to meet the needs of struggling learners.

**Policy development.** Terhart (2013) acknowledged that increasing the pressure of accountability on schools does not result in improved teaching practices. Educators are open to school improvement, in general, but when asked about personal decisions to change practice, they are averse to the concept (Terhart, 2013). Increased pressure from the federal (NCLB) and state (APPR) education departments for accountability finds teachers fearful of failing and thus resisting change (Swain, 2013). It is human nature to fear being evaluated (Bechar & Mero-Jaffe, 2013), and this is heightened when the process ranks and sorts teachers. Darling-Hammond (2013) stated that this process ultimately undermines the learning community within the schools.

The study identified collaboration with peers within their own school district as the primary impetus for changing classroom instructional practices. Similarly, Darling-Hammond (2013) stated that “90% of the nation’s teacher’s report that their colleagues contribute to their teaching effectiveness” (p. 60). The New York State Education Department needs to re-evaluate the goal of the APPR system. If its intentions are to follow Greve’s (2003) performance theory and change classroom instructional practice to improve student learning, then it needs to become a formative and engaging process. The
ranking and sorting of teachers that exists in the current system must be replaced by a formative process that encourages change. Change is the by-product of social learning (Fullan, 1991), and therefore the APPR process needs to encourage collaboration and remove competition and ranking.

Creating a formative evaluation system that incorporates professional collaborative activities to support learning would provide an atmosphere conducive to change. Teachers would change classroom instructional practice when they feel supported and are encouraged by their school district. Sebastian and Allensworth (2012) identified the correlation between professional development concentration on instruction and the enhancement of collaboration between teachers, increased opportunities for reflection, and positive feedback conversations. Taylor and Tyler (2011) furthered this idea by stating the benefits afforded teachers when they work collaboratively with effective teachers. When teachers are able to work with one another to grow as educators, they would have a positive impact on student learning.

Secondly, a formative teacher evaluation system that focusses on teacher learning is needed in order to make APPR meaningful. Benedict et al. (2013) furthered the notion of formative evaluations as a means by which teachers engage in continuous learning. Implementing periodic benchmarks by which teachers can reflect on their classroom instructional practices with peers and administrators would provide valuable feedback. Professional development focused on research-based instructional practices would provide tools and options for teachers as they strive to meet the benchmarks. The goal of evaluation is to improve practice, (Calabrese, Sherwood, Fast, & Womak, 2004; Hill &
Grossman, 2013), and formative evaluation tools would provide the feedback necessary for Greve’s (2003) performance theory to take root.

**Practice.** School districts can maximize opportunities to improve classroom instructional practice without the need for modifications to the New York State APPR law. Providing time within the school day for teachers to work collaboratively in a professional learning community requires prioritizing scheduling, and it does not incur fiscal expenditures. In accordance with performance feedback theory, once the atmosphere is conducive, people with the most working knowledge of the problem create solutions to address it. Applied to education, teachers have the best knowledge by which to improve student learning.

Providing teachers with time to reflect upon their practice with an opportunity to work with other teachers about best practice creates an atmosphere conducive to taking chances to change practice. Taylor and Tyler (2011) espoused the catalyst of peer collaboration creating an atmosphere where teachers are willing to change practice. This timeframe of collaboration and reflection creates a feedback mechanism that is the basis of performance feedback theory.

Performance feedback theory holds that individuals and organizations cyclically review and compare their performance to previous history or comparable models as a form of continuous improvement (Greve, 2003). Comparable models in education are the peers who teach within the same school district. The second primary impetus for causing teachers to change classroom instructional practice is student performance during class time instruction. Likewise, Greve (2003) stated that this time of uncertainty is where organizational performance influences decision making. When students are not

126
successful during class, teachers adapt their instruction to meet the needs of the student immediately. If students are learning at an acceptable pace, then instruction continues as planned without changes. Greve (2003) stated that uncertainty is the point at which the employee realizes there is a need for change because performance is not achieving the preset goals.

Ultimately, providing time for teachers to collaborate about what they see working or not working during classroom instructional time would provide conditions conducive to changing practice. Feedback regarding student performance serves as the catalyst for change and further reflection. This cyclical pattern, known as performance feedback theory, places the onus of improving student learning in the hands of the practitioners who work directly with the students. Unlike APPR, which costs close to $700 million annually to implement and does not cause teachers to change practice, allowing time for collaboration requires prioritizing scheduling to allow time for teachers to meet.

Adhering to the WPBA idea espoused by Guraya (2015), school districts should specifically outline the metrics on the rubrics that they prioritize as critical for student success. No more than five metrics should be highlighted in any one year. Following this identification, school district leaders need to provide training for the various levels of teachers so that the teacher has a clear definition of what the metric is, how it should look employed in the classroom, and has the ability to articulate the impact it will have on student learning. Along with training, time should be provided during the school day for collaboration among peers. This could take place within common planning periods or
combined duties. Keeping a focus on what is best for students would also include providing time for peer collaboration on staff development days.

Collaboration is not open to interpretation. Time set aside for teachers to work together would also include some type of communication with district leaders regarding what they are working on, a timeline for implementation, and corresponding updates. Time is a valuable commodity, and school districts have little to waste in this domain. Requiring teachers to be accountable for their time working together on planning, instruction, or assessment is good fiscal practice on the part of the school district.

Good performance assessment in the absence of APPR would include an evaluation system void of labels and scores. More specifically, it would include a standards-based design using the instructional and delivery components of the state-approved assessment rubrics. Three to four of their metrics would be focused on each school year. One to two metrics would be chosen by the district, and an additional one to two metrics would be tailored to meet the development needs of the teacher. These tailored metrics would be decided upon during a conference at the start of each school year. During the school year, four to six observations would take place with each teacher, followed by peer-coaching feedback sessions. Observations and peer-coaching feedback sessions are provided by district administrators or mentor teachers. Mentor teachers are senior teachers who have been tapped by their administrators because of their excellence in the classroom. Once tapped, mentor teachers would be trained alongside administrators regarding how to assess a teacher and how to serve as a peer coach.
Mentor teachers would also serve as role models for both new hires and staff in need of professional development in particular instructional areas. Mentors would discuss the instructional strategy that the teacher needs to address. The teacher would then observe a lesson where the mentor teacher skillfully employed the strategy. A debriefing session would follow where the teacher would discuss what he or she learned and articulate how he or she would employ what was learned in the classroom. This would follow with the mentor teacher observing the teacher during a lesson where he or she incorporated the strategy. Again a peer-coaching discussion would follow.

To address the time out of the classroom that is necessary for such practice, administrators could cover classes or work with the schedule to optimize staff availability. Combining study halls and other duties to free up teachers to cover classes for peer coaching is one example of how to maximize staff in order to promote staff development.

Employing this type of a performance assessment program would be formative in nature and create a collegial working environment for all faculty and staff. The evaluative element would be exchanged for a discussion about good practice and student achievement.

Lastly, through the lens of social justice, the practice of teacher evaluation needs to change in order to meet the goal of having an effective teacher in every classroom. New York State has many pockets of success and despair regarding education. Creating a formative evaluation process that supports the classroom teacher in an encouraging environment would do more than just identify strong and weak teachers. It would provide a road map for developing pedagogical skills that would improve achievement.
for every student. Unless we change the current evaluation process, the teachers are not the only people impacted. Students deserve to have effective teachers in their classrooms, and until this is the reality for all students, the disparity between successful learners and struggling learners will remain. The disparity is particularly concerning for students in urban and rural settings throughout New York State.

**Conclusion**

Research supports the notion that the instrumental factor regarding student success is the classroom teacher. Tucker and Stronge (2005) stated that effective teachers have the power to transform learning for every student. Federal and state legislative attempts to address the issue focus on holding teachers accountable for student success. Federal re-authorization of the Elementary and Secondary Education Act in 2009 led to the creation of the Race to the Top competitive grant program. This grant provided federal money to states that enacted ambitious reforms that focused on teacher evaluation. New York State passed rigorous teacher evaluation laws to qualify for federal grant money totaling $696,646,000.

While the research supports teacher evaluations as a tool to improve classroom performance on the part of the teacher, it is imperative that they do more than just identify strong and weak teachers (Papay, 2012). To be effective, teacher evaluation tools must assist teachers in assessing their performance as well as provide a road map for developing pedagogical skills that will improve student achievement. Recognizing the need for formative tools to assist teachers in continuous learning, the study explored the extent to which APPR promotes change in classroom instructional practice by teachers.
Performance feedback theory provides the theoretical framework for teacher evaluations. The theory shares basic elements that are included in teacher evaluation rubrics: (a) goal setting, (b) performance feedback, and (c) establishment of aspiration levels to change practice (Greve, 2003). Performance feedback theory is a diagnostic tool used to discover problems (Ryan, 2004), and it is a cyclical process for continuous improvement (Greve, 2003). APPR employs the same tenets of performance feedback theory with regard to identifying areas of concern for classroom teachers, and then making necessary changes to improve. Continuous improvement through feedback is synonymous with APPR and performance feedback theory. According to Greve (2003), when goals and aspiration levels are not met, decision makers are inclined to take action to improve performance. Similarly, APPR is designed to promote changes in teacher practice based upon rigorous standards-based evaluations.

The purpose of the study was to examine the extent to which public school English, math, science, and social studies teachers of Grades 9-12 are changing classroom instructional practice based upon the implementation of New York State’s Annual Personnel Performance Review. Considerable time, effort, and financial expenditure on the part of school districts are spent implementing APPR in hopes of changing teacher practice to improve student learning. Therefore, it is critical to understand the extent to which APPR is making an impact.

A sequential mixed-methods study explored the following research questions:

1. From the perspective of selected, experienced lead evaluators within Central New York school districts, who have 2 years of experience under the new
APPR process, what are the critical components of classroom instructional practice contained in the state-approved evaluation rubrics?

2. From the perspective of selected teachers, in Grades 9-12 within Central New York school districts, who have 2 years of experience under the new APPR process, what are the primary impetuses for teachers to change classroom instructional practice?

3. Given the critical components of classroom practice, as identified by question 1, from the perspective of selected English, math, science, and social studies teachers in Grades 9-12 within Central New York school districts, in which component(s) did they make the most changes in instructional practice over the past 24 months?

4. Given the changes identified by the same participants in question 3, what was the primary impetus for changes made in their classroom instructional practice?

To become an effective teacher requires purposeful work on the part of teachers and principals who work together as reflective practitioners. Principals observe teachers as part of a performance evaluation process designed to improve classroom practices and to guide overall professional development for the school district. The power of feedback to improve instruction is influenced by the instruments and procedures designed for the evaluation process (Hill & Grossman, 2013).

The significance of the study focused on the overall affect that the APPR process has in regards to improving student learning by holding teachers more accountable. Lessons from the study focused on what drives instructional changes in the classroom
along with the overall impact of the APPR process on teacher’s instructional practice. No change was reported by 20% of the participants in the study, indicating that their feedback reflected acceptable student performance. The 80% of the respondents who did indicate making changes during the 2013-2014 and 2014-2015 school years indicated that the impetus to change was due to collaboration with their peers and feedback provided by students during class time instruction. As an impetus to promote change for English, math, science, and social studies of Grades 9-12 in Central New York, APPR was identified nine out of 20.

Significant research is available that focuses on teacher evaluation and its impact on student learning. None is more poignant than Papay’s (2012) acknowledgement that teacher evaluation must do more than just identify strong and weak teachers. Teacher evaluation is not an exact science. Teachers often have difficulty accepting the evaluation process due to unclear communication, perceived biases, lack of trust in their evaluators, and perceived punitive applications of the process (Tuytens & Devos, 2013). Administrators are one variable that affects the willingness of the teacher to enact suggested changes resulting from the overall evaluation system (Danielson, 2010b; Pearce & Porter, 1986; Runhaar, Sanders, & Yang, 2010; Whiting et al., 2010). With this research in mind, it is evident that APPR may not be actualizing its initial intentions. Teachers decide to change practice based on personal aspirational levels and feedback resulting from student performance and collaboration with peers. Teachers collaborate with one another in an environment where deciding to make changes is personal and safe. They base their decisions on formative assessments made during instructional class time
with students, which is an organic approach that emphasizes the grass roots level of education.

The context for the two-phased, sequential, mixed-methods study focused on New York State public school districts in Cortland, Onondaga, and Oswego counties. Phase one consisted of two qualitative focus groups, which informed the quantitative survey for phase two. A focus group of five selected and recruited lead evaluators addressed research question 1 concerning the critical components of classroom instructional practice. Lesson planning and classroom atmosphere emerged as the two major categories of classroom instructional practice.

A second focus group comprising six teachers met to address research question 2 concerning the impetuses that cause teachers to change classroom instructional practice. The categories identified as the major impetuses to change practice included feedback, professional development and continuous learning, and planning.

Phase two of the study included a 10-question survey using Qualtrics online software that was made available to participants from September 14, 2015 through September 28, 2015. Of the 27 public school districts invited to participate in the study 18 accepted, for a participation rate of 66%. Skip logic was employed in the survey’s initial questions to pare down 292 participants to 114 English, math, science, and social studies teachers of Grades 9-12. Participation was furthered narrowed as seven teachers did not meet the minimum 2-year experience with APPR, and 21 teachers reported not making any changes during the 2013-2014 or 2014-2015 school years. Final participation in the study included 83 participants.
Lesson planning was identified as the classroom instructional component where teachers reported making the most changes, thus answering research question 3. All 83 participants in the study reported some degree of change in lesson planning.

Descriptive statistics were employed to answer research question 4: Given the changes identified by the same participants in research question 3, what was the primary impetus for changes made in their classroom instructional practice? Collaboration with peers within one’s own school district and student performance during class time instruction were the two primary impetuses that caused teachers to change classroom instructional practice. Slightly more than 50% of the participants identified the formal APPR process as an impetus for changing their practice in the classroom.

The implications of this study address the professional learning community found in schools. Teachers are changing classroom instructional practice in an effort to improve student learning. The impetus to do so is inherent in their personal commitment to continuous learning in a supportive professional learning community. Creating a collaborative atmosphere where teachers are willing to set high personal goals for student learning is more conducive to the educational process than creating a rigid and costly system that holds teachers accountable by proxy. Implementing a formative system that fosters collaboration and feedback as the evaluative tool will garner more support from teachers than continuing with a process not supported by the classroom teacher.

The study is limited in its generalizability as only 83 English, math, science, and social studies teachers of Grades 9-12 from Central New York participated. Secondly, the survey, made available in September, is a busy time for teachers as they are prepping for the new school year. Participation in the study may have been greater if it were
released in late May or June when the APPR process is active for most secondary teachers.

Recommendations included in the study include suggestions for further research, policy changes, and practice recommendations. Applying the principles of this study to elementary or middle schools would be valuable as each level offers unique pedagogical challenges and expectations. New York State identifies highly successful schools as Rewards Schools and those with significant struggles with student performance as Priority Schools. Researching to what extent APPR is making a difference in these select schools would be a valuable tool for educational leaders.

A recommendation for changing policy concerns the focus of the APPR process. Instead of employing a supervisory and accountability system, it should employ a formative development plan. By creating a formative evaluation system that incorporates professional collaborative activities that support learning, New York State would be able to create an atmosphere conducive to change. Teachers will change classroom instructional practice when they are supported and encouraged by the school district.

Recommendation for changing practice is to create time within the school day for teachers to collaborate. Whereas APPR costs close to $700 million to implement annually, prioritizing scheduling and creating a professional learning community atmosphere within the school district incurs no financial cost. School districts do not have to wait until APPR is changed, as they can improve teacher pedagogy by implementing collaboration time. The focus for the district should be on formative development of the instructional staff.
Lastly, and with great emphasis, New York State needs to refocus the tenets of the APPR system so that it metes out social justice opportunities for each teacher and student across the state. Having an effective teacher in every classroom throughout the state should not be a goal, it should be the reality. Unfortunately, as long as the APPR process remains subjective and competitive, the environment that fosters collaboration and professional development will not thrive. We owe our teachers, who have the power to transform learning for every student (Tucker & Stronge, 2005), an environment where they are supported, guided, and trusted so that they, in turn, can do the same for their students.
References


Appendix A

Email Letter to Participants Requesting Their Participation in the Lead Evaluator Focus Group

Good afternoon name of participant. I am currently working on my doctorate in Executive Leadership through St. John Fisher, and would like to collect data from lead evaluators as part of my research. In order to defend my dissertation proposal, I need to prove that I will have access to teachers in Central New York for completing the focus group portion of the dissertation study.

By replying in the affirmative to this email, you are agreeing to forward to participate in a focus group this summer. The focus group will discuss critical components of classroom practice. All information and data from the survey will be kept confidential. No individual district, nor teacher, will be identified in the presentation of the study’s findings, as all data will be aggregated.

Should you agree to participate in the focus group, an informed consent letter outlining the study and your protected rights, will be presented for your signature prior to conducting the discussion.

The purpose of the dissertation study is to examine the extent to which public school teachers are changing classroom practice based upon implementation of New York Education Law §3012-c, known more commonly as the APPR.

Thank you,

Susan T. Vickers

St. John Fisher Doctorate in Executive Leadership Student
Appendix B

Letter of Introduction to the Participants – Focus Group with Lead Evaluators

Dear Lead Evaluator:

I am a doctoral student at St. John Fisher College (SJFC) in Rochester, New York. As part of my doctoral research, I am conducting a study to determine the extent to which teachers are changing instructional practice based upon implementation of New York State’s Annual Professional Performance Review (APPR) process. For this study I am interested in hearing what you determine to be critical components of classroom instructional practice for teachers of English, math, science, and social studies in grades nine through twelve.

In this study, you will be asked to participate in a focus group discussion that should last approximately one hour. There are minor risks to you from participation in this focus group, such as the risk that a participant in the focus group will repeat another participant’s responses to a third party.

All focus group participation will be confidential; a number will be used to identify your participation when the discussion is transcribed. All notes and recording of the focus group will be locked at the researcher’s home and destroyed after three years (December 2018).

Your participation in this study is completely voluntary and you may withdraw at any time without negative consequences. If you wish to withdraw at any time during the study you may simply stop participating.

Please feel free to contact me, Susan Vickers at [redacted], if you would like to discuss anything about this study. The Institutional Review Board (IRB) of St. John Fisher College has reviewed and approved this research proposal. For any concerns regarding confidentiality, please call Jill Rathbun [redacted] She will direct your call to a member of the IRB at St. John Fisher College.

Thank you for your willingness to help with this research! Your ideas are valuable and will help determine the critical components of classroom instructional practices.

Susan

Susan Vickers
Doctoral Student and Researcher
St. John Fisher College
Doctorate in Executive Leadership
Appendix C

Informed Consent Forms for Focus Groups

Title of Study: A Sequential Mixed-Methods Study of Teacher Evaluation in Central New York: The Extent to which Annual Professional Performance Reviews Change Classroom Instructional Practice

Name of researcher: Susan Vickers, Ed.D. Candidate, SJFC

Faculty Supervisor: C. Michael Robinson, Ed.D. (crobinson@sjfc.edu)

Purpose of Study: To examine the extent to which public school teachers of social studies, science, English, and math in grades nine through twelve are changing classroom practice based upon implementation of New York Education Law §3012-c, known more commonly as APPR.

Place of study: Public school districts within Onondaga, Cortland, and Oswego counties

Length of participation: Focus groups will last between 45 – 60 minutes. A digital recorder will be used to record the conversation, and the researcher will both facilitate the discussion and take field-notes. The focus group will include 6-8 participants and the researcher. Questions will be posed by the researcher for the participants to discuss. Follow-up questions may occur based upon the discussion. The focus group will conclude after the main questions have been addressed.

Risks and benefits: There are minor risks from participation in the focus group, such as the risk that a participant in the focus group will repeat another participant’s responses to a third party. There are no benefits.

Method for protecting confidentiality/privacy: Data obtained from the focus groups will find participants identified in numeric form. All data, including the code-book, will be locked at the researcher’s home and destroyed after 3 years. Results of the study that are incorporated into publication will not utilize any identifying information.

Your rights:
As a research participant you have the right to:
• Have the purpose of the study, and the expected risks and benefits fully explained to you before you choose to participate
• Withdraw from participation at any time without penalty
• Refuse to answer a particular question without penalty
• Be informed of appropriate procedures or courses of treatment, if any, that might be advantageous to you
• Be informed of the results of the study. If you agree to participate, you must check “yes” on the consent form; a check mark next to “no” will end further communication.

If you experience emotional or physical discomfort due to participation in this study, please contact the researcher, Susan Vickers at [contact information] for appropriate referrals.

The Institutional Review Board (IRB) of St. John Fisher College has reviewed this project. For any concerns regarding confidentiality, please call Jill Rathbun [contact information]. She will direct your call to a member of the IRB at St. John Fisher College.

☐ Click here to indicate that you have read and understood the study described above and have access to a copy of this form
☐ Click here to indicate that you agree to participate in the study as outlined above
Appendix D

Email Letter to Participants Requesting Their Participation in

the Teacher Focus Group

Good afternoon name of participant. I am currently working on my doctorate in Executive Leadership through St. John Fisher, and would like to collect data from teachers as part of my research. In order to defend my dissertation proposal, I need to prove that I will have access to teachers in Central New York for completing the focus group portion of the dissertation study.

By replying in the affirmative to this email, you are agreeing to forward to participate in a focus group this summer. The focus group will discuss impetuses for changing practice in the classroom. All information and data from the survey will be kept confidential. No individual district, nor teacher, will be identified in the presentation of the study’s findings, as all data will be aggregated.

Should you agree to participate in the focus group, an informed consent letter outlining the study and your protected rights, will be presented for your signature prior to conducting the discussion.

The purpose of the dissertation study is to examine the extent to which public school teachers are changing classroom practice based upon implementation of New York Education Law §3012-c, known more commonly as the APPR.

Thank you,

Susan T. Vickers

St. John Fisher Doctorate in Executive Leadership Student
Appendix E

Letter of Introduction to the Participants – Focus Group with Teachers

Dear Teacher:

I am a doctoral student at St. John Fisher College (SJFC) in Rochester, New York. As part of my doctoral research, I am conducting a study to determine the extent to which teachers are changing instructional practice based upon implementation of New York State’s Annual Professional Performance Review (APPR) process. For this study I am interested in hearing what you determine to be impetuses for changing classroom instructional practice.

In this study, you will be asked to participate in a focus group discussion that should last approximately one hour. There are minor risks to you from participation in this focus group, such as the risk that a participant in the focus group will repeat another participant’s responses to a third party.

All focus group participation will be confidential; a number will be used to identify your participation when the discussion is transcribed. All notes and recording of the focus group will be locked at the researcher’s home and destroyed after three years (December 2018).

Your participation in this study is completely voluntary and you may withdraw at any time without negative consequences. If you wish to withdraw at any time during the study you may simply stop participating.

Please feel free to contact me, Susan Vickers at [redacted] if you would like to discuss anything about this study. The Institutional Review Board (IRB) of St. John Fisher College has reviewed and approved this research proposal. For any concerns regarding confidentiality, please call Jill Rathburn [redacted] She will direct your call to a member of the IRB at St. John Fisher College.

Thank you for your willingness to help with this research! Your ideas are valuable and will help determine the critical components of classroom instructional practices.

Susan

Susan Vickers
Doctoral Student and Researcher
St. John Fisher College
Doctorate in Executive Leadership
Email seeking District Permission to Participate in the Study

Superintendent ____________,

Good afternoon Brian. I am currently working on my doctorate in Executive Leadership through St. John Fisher, and would like to collect data from your teachers as part of my research. In order to defend my dissertation proposal, I need to prove that I will have access to teachers in Central New York for completing the survey portion of the dissertation study.

By replying in the affirmative to this email, you are agreeing to forward to your teachers a link to a survey on Qualtrics (a collegiate survey tool very similar to Survey Monkey) later this spring. The survey focuses on classroom instructional practices and what causes teachers to change practice. All information and data from the survey will be kept anonymous. No individual district, nor teacher, will be identified, and all data will be aggregated.

The purpose of the dissertation study is to examine the extent to which public school teachers are changing classroom practice based upon implementation of New York Education Law §3012-c, known more commonly as the APPR.

Thank you,

Susan T. Vickers
St. John Fisher Doctorate in Executive Leadership Student
Appendix G

Survey on Qualtrics

Q1:

Informed Consent Form

Title of Study
The Extent to which Annual Professional Performance Reviews Change Classroom Instructional Practice: A Sequential Mixed-Methods Study of Teacher Evaluation in Central New York

Name of Researcher: Susan Vickors (315)440-6839, Ed.D Candidate, SJFC

Faculty Supervisor: Dr. C. Michael Robinson, crobinson@sjfc.edu, Dissertation Chair

Purpose of study:
To examine the extent to which public school teachers of social studies, science, English, and math in grades nine through twelve are changing classroom practice based upon implementation of New York Education Law §3012-c, known more commonly as the APPR.

Place of study: Public school districts with Onon, Cort., & Oswego counties

Length of study: 10 minutes

Risks and benefits: The expected risks and benefits of participation in this study are explained below:
There are no risks nor benefits for participating in the study.

Method for protecting confidentiality/privacy:
All your responses will be anonymous. The researcher sent a link to your district to disseminate to all faculty, and therefore no email addresses have been shared. All data will be locked at the researchers home & destroyed after 3 years. Digital data will be kept on an external hard drive at the researchers home under lock and key, and destroyed in December 2018. Results of the study that are incorporated into publication will not utilize any identifying information.

Your rights: As a research participant, you have the right to: Have the purpose of the study, and the expected risks and benefits fully explained to you before you choose to participate. Withdraw from participation at any time without penalty. Refuse to answer a particular question without penalty. Be informed of appropriate procedures or courses of treatment, if any, that might be advantageous to you. Be informed of the results of the study. If you agree to participate, your continuation with the survey conveys your signed consent for participation.

If you experience emotional or physical discomfort due to participation in this study, please contact the researcher, Susan Vickors at [unredeemable] for appropriate referrals.

The Institutional Review Board (IRB) of St. John Fisher College has reviewed this project. For any concerns regarding confidentiality, please call Jill Rathbun[unredeemable]. She will direct your call to a member of the IRB at St. John Fisher College.

Questions about the Research
If you have questions regarding this study, you may contact Susan Vickors, [unredeemable], stv04448@sjfc.edu.

Questions about your Rights as Research Participants
If you have questions you do not feel comfortable asking the researcher, you may contact (Dr. Robinson), crobinson@sjfc.edu.

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

☐ yes
☐ no
Q2. Do you teach English, social studies, math, or science?

- Yes  
- No

Q3. Do you currently teach English, math, social studies or science to students in grades 9 - 12?

- Yes  
- No

Q4. Were you evaluated using the new APPR standards-based rubric system for the last two years (2013-14, and 2014-15)?

- Yes  
- No

Q5. Which subject area contains the courses you predominantly teach?

- English  
- Math  
- Social Studies  
- Science

Q6. During the 2013-14, or 2014-15 school year, have you made any changes in your classroom instructional practices and/or classroom atmosphere design?

- Yes  
- No

Q7. Have you made any changes in your classroom instructional practice over the past two years (2013-14, and 2014-15) in the area of lesson planning? Lesson planning for this survey may include the actual lesson plan formatting, as well as any or all of the following:

- design and implementation of the lesson plan
- direct instruction
- language acquisition/teaching vocabulary
- data-driven instruction
- identifying and focusing on student level of cognition
- differentiated learning
- lesson plan revision
- activities/strategies to improve student engagement
- anticipatory set
- incorporation of technology
- formative/summative assessments
- feedback to students
- learning goals
- agenda's

- Yes  
- No
Q8. Please indicate the degree to which you changed classroom instructional practices in the area of lesson planning based upon the following impetuses:

<table>
<thead>
<tr>
<th></th>
<th>Did not make changes as a result of this impetus</th>
<th>Moderate changes (tweak/slight modification)</th>
<th>Significant changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>anonymous survey/critique from students</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>formal feedback from administrators (outside of APPR)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>informal feedback from administrators (outside of APPR)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>feedback from parents</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>informal feedback from students</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>student performance on summative assessments during the year</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>student performance on formative assessments during the year</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>student performance during class time/instruction (reflection, re-evaluate, and re-assess)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>collaborate with peers outside of your own school district</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>collaborate with peers within your own school district</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>collaboration with student teachers</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>professional development offered by your own school district</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>ongoing school district initiatives</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>attendance at conferences geared towards K-12 education</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>attendance at content/level specific conferences</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>articles in professional journals or books written about educational instructional practices</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>curriculum changes within the school district</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>curriculum changes as a result of Common Core implementation</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>personal experiences as a parent</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>directly resulting from the APPR process at your school</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
Q9. Have you made any changes in your classroom instructional practice over the past two years (2013-14, and 2014-15) in the area of classroom atmosphere? Classroom atmosphere for this survey is defined as any one or all of the following options:

- the manner in which the classroom management practices are designed and implemented
- the rapport developed between student and teacher
- the relationship built between the teacher and their classes
- the relationship built amongst students
- the fluid movement to teacher as facilitator
- student's controlling the pacing and path of their learning

☐ Yes  ☐ No
Q10. Please indicate the degree to which you changed classroom instructional practices in the area of classroom atmosphere based upon the following impetuses:

<table>
<thead>
<tr>
<th>Yes: Did not make changes as a result of this impetus</th>
<th>No: Moderate changes (weak/moderate modification)</th>
<th>Total: Significant changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymous surveys/critiques from students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal feedback from administrators (outside of APPR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal feedback from administrators (outside of APPR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback from parents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal feedback from students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student performance on summative assessments during the year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student performance on formative assessments during the year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student performance during class time/instruction (reflection, re-evaluate, and re-assess)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborate with peers outside of your own school district</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborate with peers within your own school district</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaboration with student teachers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional development offered by your own school district</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongoing school district initiatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance at conferences geared towards K-12 education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance at content-level specific conferences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Articles in professional journals or books written about educational instructional practices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum changes within the school district</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum changes as a result of Common Core implementation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal experiences as a parent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directly resulting from the APPR process at your school</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H

Teacher Focus Group Form

Research: Discussion of critical components of classroom instructional practice for teachers of English, math, science, and social studies of grades 9-12 that have changed over the past two years

Date: _______________________________________

Time: _______________________________________

Location: _______________________________________

Informed Consent Forms signed? ________________________

Notes to focus group members:

➢ Thank you for your participation. I believe your input will be valuable to this research and in the continued refinement of teacher evaluation as a tool to improve student learning.

➢ Confidentiality of responses is guaranteed.

➢ Approximate length of focus group discussion is one hour with 2 main questions.

➢ Purpose of research: To understand the extent to which APPR is changing classroom instructional practice for teachers of English, math, science, and social studies in grades nine through twelve.

Focus Group Question:

1. While thinking about your classroom instructional practice, what impetus has caused you to change your practice over the past two years?

To encourage interaction, the researcher will have the option to ask follow-up or clarifying questions related to the research questions. For example, “please tell me more about…”
Appendix I

Lead Evaluator Focus Group Form

Research: Discussion of critical components of classroom instructional practice for teachers of English, math, science, and social studies of grades 9-12

Date: _______________________________________

Time: _______________________________________

Location: _______________________________________

Informed Consent Forms signed? ________________________

Notes to focus group members:

➢ Thank you for your participation. I believe your input will be valuable to this research and in the continued refinement of teacher evaluation as a tool to improve student learning.

➢ Confidentiality of responses is guaranteed.

➢ Approximate length of focus group discussion is one hour with 2 main questions.

➢ Purpose of research: To understand the extent to which APPR is changing classroom instructional practice for teachers of English, math, science, and social studies in grades nine through twelve.

Focus Group Questions:

1. While evaluating teachers of English, math, science, and social studies in grades 9-12, what are the critical components of instructional practice that you look for?

2. How are these critical components of instructional practice delineated within the NYS approved scoring rubrics?

To encourage interaction, the researcher will have the option to ask follow-up or clarifying questions related to the research questions. For example, “please tell me more about…”
Appendix J

Introductory Letter with Link to the Survey

Dear Teachers:

I would like to thank you in advance for accessing and completing the attached survey. It is a short 10 question survey that focuses on your classroom instructional practice over the past two years. Completion of this survey is anonymous, and the resulting data will be analyzed and incorporated into my dissertation study on APPR. The link to the survey is open until 9:00pm on September 28th.

Respectfully,
Susan Vickers
Doctoral Candidate at St. John Fisher College

Link to the survey:
https://sjfc.co1.qualtrics.com/SE/?SID=SV_esQ5JshKsbhtbq5