The Impact of Performance Evaluations on Teachers' Instructional Practices and Student Achievement: Evidence from a Rural School District

Lee Ann Calvert

Follow this and additional works at: http://digitalshowcase.lynchburg.edu/etd

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

Recommended Citation

The Impact of Performance Evaluations on Teachers’ Instructional Practices and Student Achievement. Evidence from a Rural School District

By

Lee Ann Calvert

DISSERTATION
Presented to the Faculty of
Lynchburg College
In Partial Fulfillment
of the Requirements for the Degree
Doctorate of Leadership Studies

Lynchburg College
March 7, 2014

Dr. Sally Selden, Chair
Dr. Roger Jones
Dr. Brian Ratliff
The Impact of Performance Evaluations on Teachers’ Instructional Practices and Student Achievement. Evidence from a Rural School District

Presented by LeeAnn Calvert and accepted on March 7th, 2014
Dedication

This dissertation is dedicated to my friends and family who have been my constant supporters and cheerleaders throughout the four years of this lengthy process. Your constant reassurance and encouragement gave me the faith and support to finish.

My husband, James Calvert, endured my horrible moods and anxiety during the process with constant support, love, and reassurance. I would like to thank him for his understanding and unconditional love.

My daughter, Lauren Calvert, is my proudest accomplishment. During this time she was understanding, loving, supportive, and proud of what I was attempting to achieve.

My parents, Nellie and Darold Fluharty, have always been my biggest cheerleaders. I would like to thank them for their unconditional love, support, and encouragement throughout this process and my life.
Abstract

THE IMPACT OF PERFORMANCE EVALUATION ON TEACHERS’ INSTRUCTIONAL PRACTICES AND STUDENT ACHIEVEMENT. EVIDENCE FROM A RURAL SCHOOL DISTRICT

LeeAnn Calvert

The purpose of this study was to investigate the performance evaluation procedures in a local county school district to find out if the performance evaluation system influenced student achievement. Two critical questions were addressed by this investigation: Does County X’s teacher evaluation system influence student achievement? Do County X teacher performance evaluations impact teacher practice? The research design included correlational and multivariate analysis to determine the relationship between teacher evaluation and student achievement. The correlation between summative performance evaluation and student performance score was found to be positive and significant $r (29) = .395, P = .034$. County X’s standard based approach that included the review of multiple sources of data produced a link between student achievement and teacher performance.

A second focus of this research investigation was to examine the impact that County X’s evaluation system had on teacher practice. Teachers were surveyed to assess the impact that the performance evaluation system has on their instructional practices. Surveys were distributed to 234 teachers who are employed in the three high schools in County X. 115 participants responded (49%). Survey consisted of 32 closed ended and
one open ended question. Closed-ended questions will be given a rating scale response choice.

The results of bivariate analysis and multiple linear regressions both showed that evaluation feedback had a significant impact on the overall quality and the overall impact of evaluation on teaching practice. Thus to improve the overall quality and impact of the evaluation process in County X, administration must improve the quality of feedback provided to teachers during the evaluation process. Sufficient time must be spent reviewing documents from documentation log during post observation conference. More time must be spent on observation and meaningful, relevant, discussions must occur during the post observation conference in order for teachers to view this process as a meaningful learning experience that promotes professional growth.
# TABLE OF CONTENTS

Dedication iv
Abstract v
Table of Contents vii
List of Tables xi

**Chapter 1: Introduction** 1
Statement of the Problem 1
Purpose of the Study 4
Significance of the Study 4
Limitations of the Study 5
Definitions 5
Research Questions 6

**Chapter 2: Literature Review** 8
Introduction 8
Past and Present Trends in Teacher Evaluation 9
Relationship between Teacher Effectiveness and Student Achievement 14
Teacher Performance Evaluations 17

  Teacher Perception of Performance Evaluation 17
    Purpose of Performance Evaluation 18
    Evaluation Improvement 18
    Effective Evaluation 20

Teacher Performance Evaluations and their Link to Student Achievement 21
Correlation Research 23
Lack of Variation in Evaluation Scores 27
  Effects of Low Stakes Evaluations 28
  Outcomes of Observation as a Single Source of Data 29
  Benefits of a Holistic Approach 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consequences of insufficient feedback</td>
<td>32</td>
</tr>
<tr>
<td>Conclusion</td>
<td>35</td>
</tr>
<tr>
<td>Summary of Research</td>
<td>35</td>
</tr>
<tr>
<td>Needed Additions to the Existing Research</td>
<td>38</td>
</tr>
<tr>
<td><strong>Chapter 3: Program Description</strong></td>
<td>40</td>
</tr>
<tr>
<td>Program Overview</td>
<td>40</td>
</tr>
<tr>
<td>Program Training/Validation</td>
<td>47</td>
</tr>
<tr>
<td>Program Evaluation Environment</td>
<td>48</td>
</tr>
<tr>
<td><strong>Chapter 4: Methodology</strong></td>
<td>50</td>
</tr>
<tr>
<td>Purpose of the Research</td>
<td>50</td>
</tr>
<tr>
<td>Research Questions</td>
<td>51</td>
</tr>
<tr>
<td>Research Hypothesis</td>
<td>51</td>
</tr>
<tr>
<td>Research Design</td>
<td>56</td>
</tr>
<tr>
<td>In-depth Study of Subset of Teacher</td>
<td>57</td>
</tr>
<tr>
<td>Sample</td>
<td>57</td>
</tr>
<tr>
<td>Measures</td>
<td>58</td>
</tr>
<tr>
<td>Variables</td>
<td>58</td>
</tr>
<tr>
<td>Summative Performance Evaluation Score</td>
<td>58</td>
</tr>
<tr>
<td>Administrators Holistic Performance Assessment Score</td>
<td>59</td>
</tr>
<tr>
<td>Student Performance Score</td>
<td>59</td>
</tr>
<tr>
<td>Summary of Measures</td>
<td>60</td>
</tr>
<tr>
<td>Analytical Plan</td>
<td>61</td>
</tr>
<tr>
<td>Survey of High School Teachers in County X</td>
<td>61</td>
</tr>
<tr>
<td>Sample</td>
<td>61</td>
</tr>
<tr>
<td>Instrument</td>
<td>62</td>
</tr>
<tr>
<td>Measures</td>
<td>63</td>
</tr>
<tr>
<td>Analytical Plan</td>
<td>63</td>
</tr>
<tr>
<td>IRB</td>
<td>64</td>
</tr>
</tbody>
</table>
Conclusion

Chapter 5: Results of the Study
Introduction
Part 1, In-Depth Study of Subset of Teachers
Sample
Data and Measurement
Analysis of Data
Examination of the Research Questions
Summary
Part 2, Survey of High School Teachers in County X
Sample
Analysis of Data
Validity
Evaluation Feedback
Multiple Sources of Data Used
Reflection on Practice
Overall Rating
Common Themes to Open Response Questions
Examination of the Research Questions
Difference in Means, Correlational and Multivariate Analysis
Summary

Chapter 6: Conclusion
Introduction
Significance of the Study
Overview of Significant Findings
Methodology
Summary of Findings by Research Question
Research Findings
Conclusions and Recommendations for County X
Limitations of the Study 144
Recommendations for Further Research 144

**References** 147

**Appendix** 154

Appendix A – Student Survey Summary 154
Appendix A – 6-12 Student Survey 155
Appendix B – Performance Standards 156
Appendix C – Teacher Survey 163
Appendix D – Holistic Performance Assessment Score 168
Appendix E – Student Performance Score 169
Appendix F – Survey Questions Mapped to Indices and Descriptive Statistics 170
Appendix G – IRB Approval 172
List of Tables

Table 3.1: County X’s Performance Rating Scale 45
Table 3.2: Logic Model of County X Evaluation System 47
Table 4.1: Teachers’ Summative Evaluations 58
Table 5.1: Descriptive Statistics for Summative Evaluation, Holistic Performance and Student Performance 72
Table 5.2: Frequency of Summative Evaluation Score 72
Table 5.3: Frequency of Holistic Performance Score 73
Table 5.4: Percent and Number of Teachers by Performance Standard 75
Table 5.5: Pearson Correlations between Teacher Summative Performance Score, Administrative Holistic Performance Assessment Score, and Student Performance Score 77
Table 5.6: Paired Sample Correlations (Seven Summative Performance Indicators and Holistic Administrative Performance Score) 83
Table 5.7: Paired Sample Statistics for Administrative Holistic Performance Assessment and Summative Evaluation Assessment 87
Table 5.8: Participant Age by Category 91
Table 5.9: Participant Teaching Experience 92
Table 5.10: Current Teaching Assignment 92
Table 5.11: Summative Evaluation Score 93
Table 5.12: Questions in Section Two of Survey Instrument -Validity 96
Table 5.13: Rating Response to Section Two Validity Questions 96
Table 5.14: Rating Response to Section Three Evaluation Feedback 98
Table 5.15: Rating Response to Section Four Multiple Sources of Data Used 100
Table 5.16: Rating Response to Section Five Reflection on Practice 102
Table 5.17: Overall Rating of the Evaluation Process 103
Table 5.18: Correlational Data to Address Research Questions 4-6 109
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Introduction

Chapter 1

Statement of the Problem

The Virginia Department of Education (VDOE) (2011) recently produced two documents addressing teacher evaluation: “Guidelines for Uniform Performance Standards & Evaluation Criteria for Teachers” and the “Virginia Standards for the Professional Practice of Teachers.” These guidelines became effective July 1, 2012, with a recommendation that school districts base 40% of teacher evaluations on student growth data. In addition to the importance placed on student performance in teacher evaluations, the VDOE (2011) has announced the Virginia Pay Performance Incentive Initiative. This initiative which was approved by the General Assembly is Governor McDonnell’s “Opportunity to Learn” education reform agenda” (VDOE, 2011, p.1). McDonnell’s three million dollar initiative awards teachers up to $5,000 for successfully increasing the achievement of student in schools that are consistently hard to staff⁶. Schools participating in this Performance Pay Initiative must use an evaluation system that is approved by the Virginia Board of Education, the evaluation must have aligned performance standards, and the evaluation must base 40% of the teacher’s evaluation on student growth (VDOE, 2011).
The federal government is also providing schools the ability to access money (59.8 million) through school improvement grants to pay for performance programs that are implemented in schools with low achievement (VDOE, 2011). Substantial funding is being provided both from the state and federal Government for school reform that links teacher pay for performance with student achievement. With this added pool of money available to schools, there is also added concern for challenges that schools will face while attempting to link teacher evaluations to student achievement gains. Steele, Hamilton, and Stecher (2010) stated in the Center for American Progress that obtaining a valid estimate of a teacher’s actual contribution to student learning is a very challenging task for school divisions. It is therefore critically important for school divisions to thoroughly investigate their performance evaluation system to insure that it is a fair and effective method to evaluate teacher performance and measure student achievement.

Tyler (2005) finds that fair performance evaluation procedures motivate employees to go beyond their daily routines and expected duties to ensure the success of the organization. Are the current procedures that schools use to evaluate teachers fair? Are these procedures effective measures of performance? Kyriakides, Demetriou, and Charalambous (2006) state that one of the major problems that education systems face is the need for an evaluation system that has a strong framework with proven methods to effectively evaluate teachers. Without such a process it is impossible for principals to accurately judge the teachers they evaluate (Kyriakides, Demetriou, & Charalambous, 2006).

Despite all of the recent mandates and incentives, many school divisions are currently without proven methods to guarantee that principals are accurately judging the
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

teachers they evaluate (Mangiante, 2011). State and federal directives have only provided school administration and central office staff with a rough outline of what evaluation components are expected with no real guidance on how to accurately measure student achievement (Mangiante, 2011). Value added assessment which is a proposed method of calculation for student achievement only works for subjects who are tested annually under a state’s accountability system. Therefore, value added assessment is only acceptably used in grades 3-8 (Steele, Hamilton, & Stecher, 2010). The proposed system of value added assessment can be easily used to compare the reading progress of students from 2nd to 3rd grade to evaluate a teacher’s impact on student achievement. Value added assessment however does not work at the high school level (Steele et al., 2010). High school classes generally test a new subject each year using an end of course Standards of Learning (SOL) tests. For example, the testing of earth science, biology, and chemistry cannot be compared to measure student growth from one year to another or from one teacher to another because the subject matter varies significantly. In addition, many subjects do not have end of the year SOL assessments. Therefore, high schools are faced with two large challenges generating valid estimates of teacher contribution to student growth and figuring out a way to include subjects that do not have annual standardized testing such as SOL tests (Steele et al., 2010). The daunting task of creating a teacher performance evaluation instrument that has a clear and accurate link between teacher performance and student achievement has been left to the discretion of school divisions. School leaders are faced with the challenge of making sure that the new instrument is valid and being implemented effectively.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

In summary, the link between teacher performance evaluations and student achievement is just starting to be established. Initiatives and incentives are being implemented to improve teacher quality, raise student achievement, and investigate the benefits of performance pay (VDOE: news, 2011). It is the responsibility of school systems to create and implement a valid instrument to measure teacher performance using student achievement data. School leaders must ensure that school systems are using an instrument for teacher evaluation that is effective, fair, and provides a crucial link between student achievement and teacher performance in order to prepare for a future that includes teacher pay based on performance and student growth data.

Purpose of the Study

To add to the emerging base of knowledge this study investigates County X’s evaluation system to find if the teacher evaluation system influences student achievement in mathematics. This performance evaluation research also investigates the impact that County X’s evaluation system has on teacher instructional practices. This investigation will be the first step in a process to determine if County X’s evaluation system is able to effectively measure teacher performance and accurately predict student achievement. It will also give critical data into the impact that the current evaluation system has on a teacher’s instructional practices.

Significance of the Study

This performance evaluation study will, however, contribute to existing research because it will be a variation of past research in three critical ways: it will include all aspects of teacher summative evaluation performance; it will include an administrative
holistic performance rating; and it will assess whether the evaluation system influences teacher practice using teacher perspective and opinions. This approach should indicate whether County X’s evaluation system is an instrument for teacher evaluation that is effective, fair, and provides a crucial link between student achievement and teacher performance.

This information will be crucial to the county being evaluated in two critical ways. First, it will be used to assess the efficiency of the current teacher evaluation practice; second, information obtained will be used to improve the current practice. This information will provide valuable insight into the need for restructuring of current procedures as well as how and what to restructure to improve the efficiency of the current practice.

Limitations of the Study

The findings of this research will be limited due to the fact that evidence is based on County X, and may not therefore be generalizable to other school districts and other subjects because of the special focus on math teachers in County X.

Definitions

The following terms are found throughout the study:

*Goal Setting*: Teacher set an annual goal, based on results of performance measures (standardized tests, benchmark assessment, or other data relevant), for improving student achievement (Carnot et al., 2007, p. 8).
Observations: Evaluators formally visit classrooms and conduct performance appraisal on the teacher, which is focused directly on the performance standards outlined by the school division (Carnot et al., 2007).

Summative Evaluation: A culminating evaluation of a teacher’s performance completed at the end of the year which assembles data from multiple formative evaluations.

Formative Evaluation: An evaluation of a teacher’s performance completed multiple times within the school year with the goal of continued improvement in performance and reflection on practice.

Multiple Sources of Data: Evaluation that considers data from multiple sources to evaluate teachers. These multiple source may include: formal observations, teacher portfolios, stakeholder surveys, achievement data, professional development, sample lesson and unit plans, etc.

Performance Standard: The major duties and responsibilities performed by the teacher (Carnot, et. al., 2007).

Performance Indicator: Examples of tangible, observable, behavior that occurs if the standard is being accomplished (Carnot, et al., 2007)

Research Questions

This research study is comprised of five chapters: a literature review; program description; methodology; results of the study; and discussion and conclusions. The literature review examines teacher evaluation, the relationship between teacher evaluation and student achievement, teacher perception of performance evaluations, and teacher performance evaluations and their link to student achievement. The next chapter
describes County X’s performance evaluation system and the program evaluation environment. The methodology chapter outlines the proposed method of research including research design, participants, and the analytical plan to interpret data collected. The next chapter is a presentation and discussion of the results of the study. Each section in this chapter will begin with description of the research sample and then follow with discussion of the analysis of data and an examination of the research questions. Finally, the last chapter is a discussion of the results of and conclusions determined from this study.

Two critical questions will be addressed by this investigation: Does County X’s teacher evaluation system influence student achievement? Do County X’s teacher performance evaluations impact teacher practice? The research design will include correlational and multivariate analysis to determine the relationship between teacher evaluation and student achievement. Teachers will be surveyed to assess the impact that the performance evaluation system has on their instructional practices. The findings of this study will be used to evaluate the effectiveness of County X’s evaluation system. This research will not only indicate whether the current system of teacher evaluation influences student achievement but it will also provide a clear picture of the impact the system has on a teacher’s classroom practice.
Chapter 2

Introduction

Chapter 2, the review of literature, investigates past and present trends in teacher evaluation from the 1700s to present day. This historic overview of teacher evaluation examines trends and criteria for evaluation that have evolved over the years with a focus on evaluation and its emphasis on instructional improvement and professional growth. In the second part of this chapter, the researcher reviews literature to establish a connection between teacher effectiveness and student achievement. The function of this section is to review the existing evidence base to assess whether teacher effectiveness is a key element for focus when addressing student achievement. The connection between teacher evaluation and student achievement as well as information obtained from the investigation of the history of teacher evaluation will establish a foundation for this study to investigate two key areas of research which include teacher perception of teacher evaluation procedures and the connection between teacher performance evaluations and student performance data.

Teacher perception of teacher evaluation procedures will provide this study with a broad perspective of how teachers view, interpret, and use data provided on performance evaluations. The research connection that is established between teacher performance evaluations and student performance will provide this study with the literature framework
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

necessary to perform an in-depth study of County X’s performance evaluation system and its link to student achievement.

Given the vast number of studies on performance evaluations, the following search terms within the time frame of 1975 through present were used: teacher perspective on teacher evaluation, teacher reaction to performance appraisals, teacher perspective on classroom observation, teacher feedback on teacher summative evaluations, teacher opinion on teacher performance appraisals, teacher performance appraisals, teacher feedback on supervision, teacher appraisal and student achievement, teacher evaluation and student achievement, appraisal of teacher performance, teacher quality and student achievement, teacher influence on student achievement, teacher effectiveness and student achievement, instructional effectiveness and student achievement, principal role in teacher evaluation, principal’s effectiveness and teacher evaluation, principal’s impact on teacher evaluation, principal’s feedback to teachers on evaluation, teacher quality and principal evaluation, teacher observation data and student achievement, teacher performance appraisals and teacher opinion.  

1

Past and Present Trends in Teacher Evaluation

In the 1700s, local government and the clergy had the power to hire and fire teachers and determine the evaluation criteria for their effectiveness. Teaching was not considered a profession at that time but more of a community service. The quality or type of evaluation varied highly across localities because there was no agreement on the importance or even the type of pedagogical expertise required (Marzano, Frontier, &

---

1 The following search engines were used: LC One Search, Education Research Complete, ERIC, ProQuest Dissertations & Theses (PQDT) and Google Scholar.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Livingston, 2011). Early evaluation was based on the wants, needs, and mores of the community and was used to determine pay increases and continued employment (Markley, 2004). Marzano, Frontier, and Livingston (2011) reported that it was not until the mid-1800s that supervision of teachers began to place emphasis on improving classroom instruction.

The early 1900s through the 1930s brought about a more scientific approach to schooling (Marzano et al., 2011). Schools were seen as factories which input raw materials (students) in an attempt to produce outputs (a productive member of society). Teacher performance was measured by productivity; productivity was measured by student learning. Student learning was measured using aptitude testing, measurable objectives, and reliable measures of student performance (Marzano et al. 2011). Using this scientific supervision approach, methods of instruction were studied, refined, and implemented to produce the highest efficiency of student learning. This data driven approach also required teachers to use prescribed strategies and behaviors in an attempt to increase the production of student learning. Performance evaluations, using the scientific supervision approach, consisted of administrators formally observing teachers and assigning a grade between A and F for the performance observed. During this time period, no attempt was made to problem solve instruction. Teacher performance was measured by student achievement of measurable objectives and graded observations (Marzano, et al. 2011).

However, it did not take long to move away from the scientific approach of manufacturing student learning and grading teachers to a focus on the teacher as an individual. Focusing on teachers as individuals began in the 1950s with Sputnik and the
Cold War, which brought additional and varied focus to education and the evaluation of school teachers (Markley, 2004). This focus was inspired by the fear that Americans would not be able to compete academically with the Soviets. This fear brought about a need to find and retain quality teachers who would prepare students to compete internationally (Markley, 2004). In this period, emphasis was placed on assisting teacher development as well as attending to teachers’ emotional needs. One important addition to teacher evaluation during this time period was the importance placed on classroom observations as a foundation for supervision and evaluation. Evaluation focused on problem solving and improvement of instruction rather than a graded exercise of a teacher’s performance (Marzano, et al., 2011).

In the early 1980s, fear once again interjected its influence into education and teacher evaluation with *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education, 1983), which declared that schools were not adequately preparing students for life, students were not learning, and students had basic skills deficits. *A Nation at Risk: The Imperative for Educational Reform* (1983) stated that improvement was needed in teacher preparation and quality. Teacher evaluation became an extension of the teacher pre-service training with a goal of continued improvement in teacher practice. Teacher evaluation focused on professional growth, where the teacher grows in competence and continues his or her educational training (Marzano, et al., 2011).

The 1980s and 1990s gave birth to the performance management movement internationally and within the United States. This movement advocated for the use of results-oriented tools to measure performance. Caillier (2010) stated that President
Clinton used performance management concepts to overhaul the federal government during his presidency. Similarly, President Bush adopted the No Child Left behind Act of 2001 (NCLB), which applied a performance management tool. This federal mandate shifted the responsibility of student achievement from the local government to state and federal governments (Caillier, 2010). NCLB contained a component to increase teacher quality by requiring every teacher to be “highly qualified” in the subject that they taught. This was accomplished by earning a degree in the subject or by passing a competency test in the content area. These mandates changed the way governments held schools accountable for student achievement and required teachers to become highly qualified in their content area, but NCLB did not mandate any changes to teacher evaluation (Washington, 2011). There appeared to be an assumption that quality teachers would produce better results as measured by school performance.

Caillier (2010) reported that critics of NCLB found that the focus on school level performance and not student level performance did not take into account teacher contributions and practice. State education agencies had been searching to find a better way to hold schools accountable for student achievement (Caillier, 2010). During the Obama administration, the focus shifted back to teacher effectiveness and a teacher’s direct impact on student achievement. The Race to the Top program (RTTT) ii initiated by the Obama administration puts pressure on school districts to redesign their evaluation system so that a significant component of teacher evaluation is based on student achievement data (Excellence in Teaching, 2010). While NCLB focused on qualified teachers, the Race to the Top (RTTT) campaign focused on the impact a teacher has on a child’s achievement. RTTT required that school districts measure teacher effectiveness
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

using multiple sources including student growth indicators as well as classroom observations (Mangiante, 2011). Mangiante (2011) criticized the proposal because clear performance criteria had not been developed. Little direction was given on how and what should be used to evaluate a teacher’s classroom performance. A gap, according to Mangiante (2011), was the existence of a research-based observation tool than could be used in conjunction with student growth data to provide a clear picture of teacher performance.

Teacher evaluation has come full circle: the current procedure to evaluate teachers is similar to the scientific approach of the 1900s where teacher evaluations were primarily based on student growth data. The pendulum swung in the opposite direction for many years, with teacher evaluation focused on instructional improvement and professional growth. Throughout history the shift of the pendulum seems to be a replacement for a research-based method to effectively evaluate teachers based on classroom observation and student growth. In the 1700s the quality and type of evaluation was highly varied because of the lack of agreement on the importance or even the type of pedagogical expertise (Marzano et al. 2011). Today little direction has yet to be given on how and what should be used to evaluate a teacher’s classroom performance. Furthermore, Marzano, et al., (2011) indicated that the knowledge base developed on the history of supervision and evaluation should not be used as a prescription for teacher evaluation. As this literature review indicates, throughout history there have been significant shifts of the pendulum in regard to teacher evaluation and supervision; past procedures do not provide a clear outline of how to effectively evaluate teachers. This
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

investigation of past and present trends in teacher evaluation, however, provides a clear understanding of the progress and development of the current evaluation system.

**Relationship between Teacher Effectiveness and Student Achievement**

Research supports that teacher effectiveness is strongly linked to student achievement (Briggs, Davis, & Cheney, 2012; Borman & Kimball, 2005; Heneman & Milankowski, 2011; Kane, Taylor, Tyler & Wooten, 2010; Mangiante, 2011; Stronge, Ward & Grant, 2011). The quantitative, non-experimental, study conducted by Kane et al. (2010) found conclusive evidence that teacher practices significantly influenced higher student achievement. Kane et al.’s (2010) study examined existing teacher observation and standardized test score data of 101 reading and 99 math teachers in Cincinnati Public Schools. They found that a student performing at the 50th percentile in the beginning of the year improved three percentage points in reading and two percentage points in math if placed with a teacher whose performance was in the top quartile (Kane et al., 2010). Similarly, Briggs, et al. (2012) found a significant link between student performance and teacher effectiveness when they compared performance of students who were assigned to an effective teacher for three consecutive years to students who were assigned to ineffective teachers for the same three years. Researchers found that students assigned to the effective teacher outperformed their comparison group by 50 percentage points (Briggs, et al., 2012).

Similar results were found in a study of 307 fifth grade teachers from three public school districts in the southeastern United States (Stronge et al., 2011). Stronge et al. (2011) examined the impact that individual teachers had on student achievement using a
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

regression based methodology, hierarchical linear modeling, to estimate student growth based on standardized test scores of students in reading and mathematics. This investigation examined how teachers with high student growth differed from teachers with low student growth. The purpose of the study was two-fold: to examine the impact teachers had on student learning and to examine the behaviors of effective (high student growth) teachers.

Stronge et al.’s (2011) research supported the findings of Kane et al. (2010) and Briggs et al. (2012). Students taught by effective teachers according to Stronge et al. (2011) out performed students taught by ineffective teachers by more than 30 percentile points in reading and mathematics. It is important to note that these three studies, Stronge et al. (2011), Kane et al. (2010), and Briggs et al. (2012), measured teacher effectiveness using measures of student growth based on performance on standardized tests.

The idea of using student growth to measure teacher effectiveness was further supported by Brophy’s (1986) meta-analysis of research linking teacher behavior to student achievement from 1950 to 1986. The findings of the research summarized were derived from studies conducted on elementary and secondary classes. Data were analyzed to find relationships between teacher behaviors and student achievement measured by standardized test scores. Brophy (1986) explored both the quantity (the degree to which teachers cover content, pace at which teachers move through the content, and the amount teachers actively engage their students) and quality (how well instructional tasks are performed) in this meta-analysis.
Brophy (1986) concluded that to improve student achievement school districts should focus on effectively measuring teachers’ classroom behaviors. His research summary of observed teacher behaviors indicated that a teacher’s classroom behavior has a direct effect on student achievement. Brophy (1986) refuted the belief that anyone can teach. He pointed out that some adults may be able to survive in a classroom setting but mere survival did not produce student achievement results. To successfully produce student achievement gains, teachers must not only have the knowledge of their subject matter but must also possess motivation and a set of pedagogical skills (Brophy, 1986). Brophy (1986) further suggested that an evaluation system that makes connections between classroom instruction and student learning gains will more effectively align teaching with student achievement.

The research findings are clear. Teacher effectiveness influences student achievement (e.g., Heneman and Milankowski 2011; Mangiante, 2011). The literature review indicates that a strong relationship exists between teacher effectiveness and student achievement. It is also clear from the literature reviewed that teacher effectiveness is a key element associated with student achievement. No conflicting studies or opposing views were found in this review of literature to argue that teacher effectiveness was not associated with student achievement. However, controversy does exist on how teacher effectiveness should be judged. Scholars and policy makers dispute whether teacher effectiveness should be judged on qualifications, instructional practices, effect on student learning, or a combination of these assessments (Stronge, et al. 2011; Lewis, Parsad, Carey, Bartfai, Farris, & Smerdon, 1999).
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

**Teacher Performance Evaluations**

The literature review establishes that teachers influence student achievement; however and as noted previously, limited information exists that gives direction on how best to measure teacher performance based on student achievement. Scholars cannot seem to agree on the methods that should be used to evaluate teacher effectiveness (Stronge, et al. 2011; Lewis, et al. 1999). While abundant sources are available to describe the qualities that effective teachers possess, there is limited information found to indicate how to implement or measure these skills. Therefore, to explore how to effectively measure teacher performance, this literature review examines two bodies of work related to teacher performance evaluations: studies exploring teacher perception of performance evaluations, and research that examines the relationship between teacher performance evaluations and student achievement.

**Teacher perceptions of performance evaluation.** An important step in the investigation of teacher performance evaluations is to examine research conducted on teacher’s perception of the evaluation process. This literature review reveals that there is limited research conducted to assess teachers’ perceptions of the evaluation process and how it affects classroom instruction. Those studies that have been conducted deliver a set of consistent findings (Derrington, 2011; Feeney, 2007; Henson & Hall, 1993; Kennedy, 2012; Kyriakides, Demetriou, & Charalambous, 2006; Mahar & Strobert, 2011; Peterson & Comeaux, 1990; Rothberg & Fenner, 1991; Tyler, 2005). The review of the literature on teacher perceptions focuses on three areas: what is the purpose of performance evaluations, where is improvement needed in performance evaluations, and improving performance evaluation effectiveness from the perspective of the teacher.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

**Purpose of performance evaluations.** Studies conducted on teacher perceptions of performance evaluations indicate that the primary purposes of performance evaluation should be for reflection and professional growth of the teacher (Peterson & Comeaux, 1990; Feeney, 2007; Rothberg & Fenner, 1991). Peterson and Comeaux (1990) interviewed 48 high school teachers in Wisconsin and Florida about their performance evaluation experiences. Teachers felt that the primary purpose of performance evaluation should be for their own professional growth. Evaluation should facilitate a teacher’s reflection on his or her classroom practice, which should in turn improve a teacher’s classroom skills and promote his or her professional growth (Peterson & Comeaux, 1990).

However, when asked about their actual experiences, the teachers interviewed acknowledged that their expectations were not met (Peterson & Comeaux, 1990). Forty percent of the teachers interviewed reported that evaluation did not cause them to reflect on their teaching (Peterson & Comeaux, 1990). Similarly, research conducted by Kennedy (2012) Derrington (2011), Henson and Hall (1993), and Rothberg and Fenner (1991) indicated that teacher performance evaluation did not facilitate reflection in practice. While teachers understand that evaluation is part of the job requirement, the existing research suggests that evaluation does not increase teachers’ professional growth or assist in the improvement of teaching practice (Derrington, 2011; Henson & Hall, 1993; Kennedy, 2012; Rothberg & Fenner, 1991).

**Evaluation improvement.** One of the major areas of improvement, as specified by research, is that teachers want more feedback and follow up during the evaluation process to grow professionally (Rothberg & Fenner, 1991; Kennedy, 2012; Mahar & Strobert,
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

2011; Peterson & Comeaux, 1990). When Rothberg and Fenner (1991) surveyed 230 teachers from central Florida, they found that teachers sought more feedback about their performance. Similarly, Kennedy’s (2012) ethnographic interview found that evaluations were not used to assist teachers and help them develop. Face to face conferences that contained feedback and follow up were needed for teachers to reflect on their practice and grow professionally (Kennedy, 2012). This is a systematic challenge as researchers have shown that quality feedback is often a missing component from the teacher evaluation process (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011).

Research revealed that teachers find evaluations that used multiple sources of data collection such as documentation logs, peer observations, stakeholder surveys, and student achievement data, along with multiple classroom observations were more helpful in providing the desired feedback to improve teacher instruction and professional growth (Peterson & Comeaux, 1990; Mahar & Strobert, 2010; Rothberg & Fenner, 1991). Mahar and Strobert (2010) surveyed teachers to compare feedback given on traditional evaluation (observation only) to multiple source evaluations to determine the quality of feedback that the teacher received. The multiple source approach to evaluation provided teachers more feedback, promoted reflection, and identified areas of need for professional growth to a larger extent than did the single source approach of observation alone (Mahar & Strobert, 2010). Compared to traditional feedback, teachers believed that feedback from multiple sources was significantly more helpful in facilitating their professional development (Mahar & Strobert, 2010). Additionally, teachers believed that feedback that is more directly linked to student progress and achievement has a greater impact on their improvement and professional growth (Mahar & Strobert, 2010).
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

**Effective evaluation.** Research indicated, based on teacher opinion, that evaluation instruments were effective only if they were perceived as fair and valid (Peterson & Comeaux, 1990; Kyriakides, Demetriou, & Charalambous, 2006; Tyler, 2005). Peterson & Comeaux (1990) interviewed high school teachers in four schools in two districts in Florida and Wisconsin and found that teacher buy-in was an essential component to effective performance evaluations. Teachers may view performance evaluations as an opportunity for interaction and growth or negatively as a source of frustration. When teachers did not view the performance evaluation instrument as a valid tool for professional growth their perception became that they only needed to pass the teacher performance evaluation, not actually learn and improve from the experience. Often times the needed component in designing an effective teacher evaluation instrument is the acceptance of the teacher being evaluated (Peterson & Comeaux, 1990).

Kyriakides’ et al. (2006) research, which used a stratified sample to survey 237 teachers, supports Peterson and Comeaux (1990) in stating that teachers should be a part of the process that generates evaluation criteria in order for them to accept the instrument as a means of judging their professional effectiveness. Teachers also believed that it was important for performance evaluation instruments to be reviewed for systematic validity and consistency among observers (Kennedy, 2012; Kyriakides et al., 2006). Teacher involvement and participation is a critical piece to ensuring that the performance instruments are considered to be fair and valid by teachers, thus increasing their effectiveness (Peterson & Comeaux, 1990).

In summary, this review has shown that teachers perceive that quality feedback and follow up are necessary components of the teacher evaluation process (Rothberg &
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Fenner, 1991; Kennedy, 2012), and feedback from multiple sources is important to improve teacher instruction (Peterson & Comeaux, 1990; Mahar & Strobert, 2010; Rothberg & Fenner, 1991). However, feedback is often a missing component in existing evaluation systems (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011).

Teachers believe a primary purpose of evaluation is to facilitate their professional growth (Peterson & Comeaux, 1990; Feeney, 2007). However, the research does not support evaluations as reflective in practice or as resulting in teacher’s professional growth (Henson & Hall, 1993; Rothberg & Fenner, 1991; Kennedy, 2012; Derrington, 2011).

The feedback that is missing from teacher performance evaluation may be the element that facilitates reflection and professional growth.

This literature review demonstrates that teacher opinion is an important piece of the evaluation puzzle that has and should be explored. In light of this finding, school districts need to examine whether teachers view the evaluation instrument as a valid measurement of their performance.

Teacher performance evaluations and their link to student achievement. The literature review shows a strong link between teacher effectiveness and student achievement (Briggs et al., 2012; Borman & Kimball, 2005; Heneman & Milankowski, 2011; Kane, Taylor, Tyler & Wooten, 2010; Mangiante, 2011; Stronge et al., 2011). The fact that teachers impact student achievement is not disputed in the literature. Strong experimental evidence exists which demonstrates that one teacher can be effectively compared to another teacher in a research environment to determine which teacher has a greater impact on student achievement Briggs et al., 2012; Borman & Kimball, 2005; Heneman & Milankowski, 2011; Kane, Taylor, Tyler & Wooten, 2010; Mangiante,
2011). So while there is no dispute over the fact that one teacher is more effective than another, there is much debate over how to measure teacher effectiveness efficiently for performance evaluations (Stronge, et al., 2011).

This section explores the connection between teacher performance evaluations and student achievement. Research in this section reveals a weak connection between teacher performance evaluation scores and student achievement, which suggests that performance evaluation systems may not adequately measure teacher effectiveness. Research studies investigating the link between teacher performance evaluations and student achievement have produced weak links but these studies often use non-experimental designs (Kane et al., 2010; Kimball et al., 2004; Milanowski, 2004; Washington, 2011).

Although findings have produced weak correlations, the results of the studies have been consistent. The results have shown evaluation systems that use a standards based approach that include multiple sources of data, produce a more significant link between student achievement and teacher performance (Hinchey, 2010; Jacob & Lefgren, 2008; Kane et al., 2010; Kimball et al., 2004; Stonge et al., 2011). The findings also reveal that one problem contributing to the lack luster findings is the lack of variation in teacher evaluation scores (Stronge et al., 2011; Kane et al., 2010; Jacob & Lefgren, 2008; Washington, 2011). Kane et al. (2010) and Washington (2011) found that the vast majority of teachers were labeled satisfactory/proficient causing there to be an insufficient range to produce correlation coefficients between performance and achievement because of the lack of variance in the data. These findings of weak correlational values, benefits of standard based approach that uses multiple sources of
data, and the lack of variation in evaluation scores will be further discussed in this section.

**Correlational research.** While many researchers have found links between teacher evaluation and student achievement, the relationships have been weak and have often used non-experimental designs (Kane et al., 2010; Kimball et al., 2004; Milanowski, 2004; Washington, 2011). This section will compare and contrast existing correlational studies, from 2000 to the present, to probe into the problem and possible reasons for lack of variation and these low correlational results.

Milanowski, in “The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence from Cincinnati” (2004), reviewed 212 teacher evaluation scores and correlated them to the reading, math, and science test scores of student in grades three through eight. This study used a quantitative, non-experimental design to investigate Cincinnati Public Schools (CPS), which had a performance evaluation system that was based on performance standards developed from Framework for Teaching (Danielson, 1996). The performance system included 16 performance standards grouped in four domains: planning and preparation, creating an environment for learning, teaching for learning, and professionalism (Milanowski, 2004). Each standard contained behavior rating scales designed in rubrics for each of the four possible levels of performance, which included: unsatisfactory, basic, proficient, and distinguished. Data were collected from observations and a portfolio prepared by the teacher which contained artifacts such as lesson plans, parent contact logs, documentation of professional development, and student work (Milanowski, 2004).
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Using this standards based evaluation system Milanowski (2004) identified teachers with higher levels of student achievement to a degree greater than chance. Milanowski (2004) stated that the empirical results from this study indicated that “evaluations produced by a relatively rigorous, standards-based system are related to an acceptable measure of student learning” (p. 49). Milanowski (2004) indicated that prior research investigations found teacher evaluations to be superficial and too simplistic to actually provide a link with student learning.

Milanowski (2004) reported that the correlational findings from this study, however, were still relatively low, with a .27 correlation for science, .32 for reading, and .43 for mathematics. He stated that high correlations may not be found due to the errors in measuring student and teacher performance and curriculum alignment with standardized tests. Even still, Milanowski (2004) linked the successful outcomes of this study to rigorous evaluations that were standards based, and the implications were clear.

Similarly, Washoe County School District evaluation system investigated by Kimball et al. (2004) was also derived from the Framework for Teaching (Danielson, 1996) and included 23 performance standards in four domains (planning and preparation, creating an environment for learning, teaching for learning, and professionalism) with behavioral descriptions contained in a four level rubric. The investigation conducted in Washoe County School District by Kimball et al. (2004) used a quantitative pretest posttest non experimental design to explore the relationship between teacher evaluation scores and student achievement. The standardized test scores of 1858 third, 1752 fourth, and 2073 fifth grade students were correlated with teacher evaluation scores. The results
from this study produced a positive correlation between teacher evaluation scores and student achievement, but it was not statistically significant.

Unlike Milanowski (2004), who included all 16 standards of the teacher evaluation instrument into the research, Kimball et al. (2004) only included 7 of the 23 teacher performance standards into the research study. The standards that were chosen by Kimball et al. (2004) for the investigation include those that were most closely associated with student achievement. Kimball et al. (2004) concluded that some important part of evaluation must be missing from the study with this limited representation of performance standards. Additionally, Kimball et al. (2004) found that both marginal and high performing teachers received the same satisfactory ratings on their performance evaluations. Kimball et al. (2004) attributed the similarity in evaluation scores to the lack of performance rubrics present in the evaluation system.

Another significant difference between the two studies was that Washoe County School District used a more typical evaluation process designed for low stakes purposes rather than to make salary determinations as the Cincinnati Public Schools did. Kimball et al. (2004) also decided that because of the low stakes effects of these evaluations, evaluators were less focused on differentiating between teachers and more focused on growth, praise, and positive morale.

Similar to Kimball et al. (2004), Kane et al. (2010) investigated a performance evaluation system that was based on a Framework for Teaching (Danielson, 1996) and selected only those standards on the teacher evaluation instrument that had a direct link to classroom practices. Kane et al. (2010) used a two-fold approach in the investigation of
99 math and 101 reading teachers in a quantitative non-experimental study. First, Kane et al. (2010) examined only the eight standards outlined in two domains (creating an environment for learning and teaching for learning) that specifically dealt with teacher practices observed in the classroom. Second, Kane et al. (2010) looked at teacher performance across all four domains and 16 standards measured on the teacher performance evaluation instrument including examination of artifacts contained in the teacher portfolio which pooled multiple sources of data. Kane et al. (2010) found a stronger correlation when all four standards were examined.

Kane et al. (2010) found significantly higher correlations in math: 0.22 (all four domains considered), 0.13 (classroom practices only), and reading: 0.21 (all four domains considered), 0.06 (classroom practices only). Researchers Kane et al. (2010) concluded from these results that it was essential for all domains to be included to accurately predict student achievement. This included those domains that were not directly recognizable during classroom observation. Teacher evaluation systems must use multiple measures not only those directly observable during classroom instruction to more effectively predict student achievement using teacher evaluation (Kane et al., 2010).

Additionally, Kane et al. (2010) acknowledged that the lack of differentiation between teacher performance scores had an effect on study results. However, Kane et al. (2010) recognized that the problem of heterogeneity in performance ratings was exposed due to the use of standards based performance ratings.

The most recent correlation study conducted by Washington (2011) attempted to correlate teacher performance scores obtained from the Performance Assessment System.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

for Teachers (PAS-T) with student achievement in reading and math. PAS-T contained eight performance standards measured using a rubric with four performance ratings (exemplary, proficient, needs improvement, and unsatisfactory). This study used a quantitative non-experimental pretest posttest model using data from 45 elementary schools. Correlational coefficient scores could not be calculated because there was not enough variance in the teacher evaluation score. Even though the study conducted by Washington (2011) was not able to successful correlate teacher evaluation data with student achievement data, it does provide a crucial piece of evidence that uncovers heterogeneity or lack of differentiation in performance ratings.

In summary, the correlational studies discussed in this section show that better correlational results are obtained when multiple sources of data were used instead of using only the data that is directly observable in a classroom environment (Kane et al. 2010; Kimball et al. 2004). The review of these studies has also brought to light the lack of variation in teacher performance evaluation scores that limit the value of correlational studies (Washington, 2011). Kimball et al. (2004) brought further clarity by suggesting that this lack of variation was due to the low stakes effects and the fact that evaluators are less focused on differentiating teacher performance and more focused on growth, praise, and positive morale. This lack of variation in teacher performance scores will be further explored in the next section.

**Lack of variation in evaluation scores.** A lack of variance is consistently reported in research linking teacher performance ratings to student achievement (Kane et al., 2010; Kimball et al., 2004; Washington, 2011). Studies have found that evaluators were not able to successfully differentiate teacher performance, and the majority of teachers
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

received the same satisfactory rating on their performance evaluations (Stronge et al., 2011; Strong, Gargani & Hacifaziloglu, 2011; Kimball et al., 2004; Kane et al., 2010; Washington, 2011). This lack of variation in performance evaluation scores and the explanations provided by the literature for this lack of variation will be discussed in this section. This section will include the effects of low stakes evaluations, outcomes of observation as a single source of data, consequences of insufficient feedback, and benefits of holistic approach.

*Effects of low stakes evaluation.* As discussed previously in the review of correlational research, many researchers attributed the lack of significant correlation between teacher performance evaluation scores and student achievement to the lack of variation in teacher performance ratings (Washington, 2011; Kane et al., 2010; Kimball et al., 2004). The investigation of Washoe County School District conducted by Kimball et al. (2004) found a positive correlation between teacher evaluation scores and student achievement, but it was not statistically significant due to the lack of variation in performance ratings. Kimball et al. (2004) decided that this lack of differentiation in performance ratings was caused by the low stakes effects of the evaluations since evaluation did not determine salary or promotion. Kimball et al. (2004) found that evaluators were less focused on differentiating between teachers and more focused on growth, praise, and positive morale.

Similarly, Washington (2010) was unable to correlate the performance evaluation scores of veteran teachers to student performance because of the lack of variation that existed in teacher performance scores. Washington’s (2010) investigation, which only included veteran teachers, showed that almost all teachers were rated in the highest
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

performance levels exemplary or proficient. Teacher performance evaluations investigated by Washington (2010) were considered low stakes since the results of the evaluation did not determine salary or promotion. By including only veteran teachers, those on continuing contract status, Washington’s study went a step further on the low stakes spectrum than the study investigated by Kimball (2004) whose study included probationary teachers.

Additionally, Kane et al., (2010), who investigated Cincinnati Public School evaluation system, acknowledged that the lack of differentiation in teacher performance scores had an effect on the correlational results obtained from linking teacher evaluation scores with student performance data. This study illuminates the fact that there must be other contributing factors besides the low stakes effect of performance evaluations since the investigation conducted by Kane et al., (2010) used a high stakes evaluation instrument that was used for salary determinations.

**Outcomes of observation as a single source of data.** While research suggests that reasons for lack of variance can be attributed to low stakes effects and that evaluators focus on growth, praise, and positive morale, there is also evidence to suggest that a single observation cannot be successfully used to differentiate performance (Kane et al., 2010; Kimball et al., 2004; Stronge, Ward, & Grant 2011; Strong, Gargani & Hacifaziloglu 2011). Kimball et al. (2004) and Kane et al. (2010) investigated a performance evaluation system that selected only those standards on the teacher evaluation instrument that had a direct link to classroom practices measured through a single source of observation. Kimball et al. (2004) and Kane et al. (2010) were unable to find significant correlations linking teacher evaluation with student achievement using
performance standards that were only measured by a single observation. Kane et al. (2010) found that observation taken as a single source of data was not able to predict student achievement better than subjective ratings.

Further investigation into a single source of observation used to predict student achievement found the quantitative analysis of 307 fifth grade teachers whose performance evaluation data taken from a single source of classroom observation was compared to student performance data (Stronge et al., 2011). This quantitative analysis was unable to distinguish between the performance of the bottom and top quartile of teachers (Stronge, et al., 2011). This study provides additional evidence that a single source of observation is not able to differentiate teacher performance.

The strongest evidence, based on experimental study design, to confirm evaluators’ inability to differentiate from a single source of observation comes from the research conducted by Strong, Gargani and Hacifaziloglu (2011). In this research project, 100 judges selected seven video clips randomly to view and rate as above average or below average. Judges consisted of administrators, teachers, teacher educators, teacher mentors, education professors, parents, undergraduate students, students, and adults with no educational experience. Findings indicated that regardless of experience, judges were unable to identify successful teachers based on video observation alone (Stong, Gargani, and Hacifaziloglu, 2011).

The research conducted by Kane et al., 2010; Kimball et al., 2004; Stronge et al., 2011; Strong, Gargani & Hacifaziloglu (2011) indicated that a single source of
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

observation is not sufficient to effectively evaluate teachers if the goal is to differentiate teacher performance and link performance evaluation to student achievement.

**Benefits of a Holistic Approach.** Differentiation of teacher performance, however, was obtained for the top and bottom quartile teachers in the research conducted by Jacob and Lefgren (2008). Jacob and Lefgren’s (2008) research study differed from existing research because they compared principals’ subjective or holistic performance ratings to student achievement data. All other research examined in this literature review used a standard based performance instrument to access teacher performance which was compared to student achievement data. Jacob and Lefgren (2008) surveyed principals and ask them to rate teachers on a scale of 1 (inadequate) to 10 (exceptional) in a variety of dimensions such as dedication, work ethic, classroom management, parent satisfaction, ability to raise reading and math achievement, and positive relationships with administrators. The sample consisted of 201 teachers in grades two through six who were teaching a core subject (math, science, history, or reading). Administrator responses were kept confidential and administrators were assured that scores would not be shared with teachers or other school personnel (Jacob & Lefgren, 2008).

Consequently, the results were different than studies that correlated standard based performance data with student achievement data. Jacob and Lefgren (2008) reported that considerable variation existed between teacher’s subjective performance ratings within schools. The results also indicated that principals were able to differentiate successfully teachers in the top and bottom categories. They, however, were significantly less successful differentiating teachers in the middle of the ability distribution. By allowing school administration to provide a holistic rating of the effectiveness of their
teachers, instead of using standards based performance instrument, Jacob and Lefgren (2008) were able to produce results that linked teacher performance evaluation to students’ achievement without the problem of lack of variation in teacher performance scores.

Similarly, when Kane et al. (2010) included all teacher standards of performance into the teacher evaluation score, a higher correlational result was found. This complete set of standards included those not directly measured by classroom observation, such as teacher portfolio, professional development, student achievement data, and teacher survey information. By including those standards measured outside of the classroom the evaluator was able to view a more complete or holistic picture of the teacher’s performance (Kane et al., 2010) which resulted in a higher correlation with student achievement data. These studies add additional support for a more holistic approach to evaluation that provides a complete picture of the teacher both inside and outside of the classroom.

**Consequence of insufficient feedback.** One of the major reasons attributed to the problem of lack of differentiation in performance evaluations is the inability or unwillingness of evaluators to give strong negative feedback when necessary (Hinchey, 2010). This section explores the importance of verbal and written feedback to the professional growth of teachers, the reasons attributed to the lack of feedback, and the link between lack of feedback and differentiation of performance evaluation scores.

A research brief produced by Gary Marx in the Principal’s Partnership (2007) reported that the most important factor in changing teacher behavior was evaluator
feedback. Likewise, Rothberg and Fenner (1991) investigated teacher perceptions of teacher assessment and found that the most frequent request by teachers was to receive more verbal and written feedback from evaluators. Without meaningful and objective feedback given in conjunction with regular reports on performance, a teacher will not grow professionally (Feeney, 2007).

Even though quality feedback appears to be a key ingredient in teacher evaluations, it is often a missing component. Blumberg (1976) reported that only 1% of the 11.5 hours of recorded teacher evaluation conferences were spent problem-solving ideas about how to improve the classroom. Additionally, only 1% of the time was spent with the teacher asking the evaluator any type of question. The collection of taped recorded evaluation conferences collected by Blumberg (1976) described conference behavior as distant, ritualistic, tense, and non-authentic. Similarly, Mahar and Strobert (2010) reported 34 years after Blumberg (1976) that respondents indicated that evaluator feedback provided “little guidance toward instructional improvement” and was found to be “vague” (p. 152). Patterson, Grenny, McMillan, and Switzler, authors of Crucial Conversations (2012), say that people use many tactics to avoid touchy issues and difficult conversations. It is unknown whether the demands of an administrator’s job are forcing them to use these ritualistic and vague conversations during teacher evaluation or whether it is simply a tactic used by administration to avoid difficult conversations.

Another hindrance to quality feedback, documented by Derrington (2011), was that most of the time not only do teachers receive similar marks, but the comments provided on their summative evaluations are “narrative phrases, resulting in strikingly similar comments for each recipient’s evaluation, causing teachers to feel that the reports
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

were a product of a cut and paste activity” (Derrington, 2011, p. 51). Additionally, the ratings associated with the remarks on a summative evaluation are simplistic and range from satisfactory to needs improvement (Feeney, 2007). Teachers who received a satisfactory rating on summative evaluation did not perceive this information as useful to let them know how they are actually performing in the classroom. It did not provide them with helpful information nor did it provide them with any motivation for improvement. The combined effect of simplistic ratings and canned responses that were devoid of any meaning reduced teaching aptitude and the connection to student learning over time (Finney, 2007). Derrington (2011) suggests that principals need to stop viewing evaluation as teachers passively receiving their judgment but as a guided activity of thinking and reflection that facilitates improvement and professional growth.

Another obstacle that inhibits principals from giving specific feedback and differentiating performance is that principals are not trained in each discipline they evaluate. Rothberg and Fenner (1991) surveyed 230 teachers and concluded that teachers did not feel that principals were effectively trained in the discipline that they were evaluating. Similarly, Marx (2007) reported that feedback from principals at the high school level was often difficult due to the expertise needed in each subject area.

There are conflicting opinions in the research as far as who and what is to blame for the lack of differentiation in the teacher evaluation process. The lack of training in subject area knowledge and the unwillingness to give quality feedback are both contributing factors to the lack of differentiation that have been observed. The research does consistently agree that quality feedback and differentiation are essential components for professional growth and reflection to occur.
As discussed previously in the review of correlational research, some researchers attribute the lack of significant correlation between teacher performance and student achievement to the lack of variance, due to the inability to differentiate among teacher performance using observation, as a common problem that exists. It is also a problem that provides a link to the reasons why correlations between teacher performance and student achievement have been weak or non-existent. A common thread that this literature review has provided is that studies that use a more holistic approach, those studies including all performance standards assessing teachers through multiple sources of data or studies that include an overall administrative view of performance, have been more successful at differentiating performance between teachers and producing better correlational results between teacher performance and student achievement. In other words, a combination approach to measuring teacher performance is superior to a single method approach (Martinez-Rizo, 2012).

Additionally, there is evidence to suggest that low stakes effects of performance evaluations (Washington 2011& Kimball et al. 2004 ) and the inability or unwillingness for administration to give quality feedback (Derrington, 2011; Hinchey, 2010; Blumberg, 1976 ) are contributing factors to the lack of variance or differentiation in teacher performance evaluations.

**Conclusion**

**Summary of research.** One key observation that was found through this literature review is that teacher effectiveness influences student achievement (Heneman
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

and Milankowski 2011; Mangiante, 2011). There is no conflicting work or opposing views to dispute that teachers have a strong impact on student achievement.

Additionally, the results from studies based on teacher perception reveal that feedback and follow up are necessary components to the teacher evaluation process (Rotheberg& Fenner, 1991: Kennedy, 2012). Feedback from multiple sources improves teacher instruction (Peterson& Comeaux, 1990; Mahar& Strobert, 2010; Rothberg & Fenner, 1991), feedback is a missing component in teacher performance evaluations (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011), evaluation’s primary purpose should be professional growth (Peterson & Comeaux, 1990; Feeney, 2007), and evaluations do not facilitate reflection or professional growth (Henson & Hall, 1993; Rothberg & Fenner, 1991; Kennedy, 2012; Derrington, 2011). Teacher opinion is a critical piece of teacher performance evaluation puzzle that must be explored by school districts. In order for school districts to facilitate effective performance evaluations, they must have the knowledge of how teachers perceive performance evaluations and whether they view them as a valid measure of their performance.

Although research found in this literature review linking teacher evaluation to student achievement has been weak and often non-experimental, the research does conclude that stronger correlations between teacher performance evaluations and student achievement are found when a more holistic approach is used that includes all of the teacher performance domains as well as multiple sources of data (Jacob & Lefgren, 2008; Kane et al. 2010; Kimball et al. 2004).
Another critical piece discussed in the literature review is the lack of variance in teacher performance ratings (Stronge et al., 2011; Kane et al. 2010; Washington, 2011). The lack of variance is a common problem that exists in teacher performance evaluations which is often due to the inability to differentiate teacher performance using observation. This problem of lack of variation in teacher performance evaluations provides a link to why past research has found only weak or non-existent correlations between teacher performance and student achievement.

Although there are conflicting views in the research to explain the lack of variance observed, inability or unwillingness of evaluators to give strong negative feedback, low stakes effects of performance evaluations, and using observation as a single source to measure performance are contributing factors (Hinchey, 2010). The research does consistently agree that quality feedback is essential for professional growth and reflection to occur.

The literature review revealed that many researchers have worked diligently for years to define practices that link teacher effectiveness data with student achievement data. It is not news in education that an ineffective teacher can have a negative impact on student outcome measures. However, despite the impact of teacher effectiveness, many divisions are still using simplistic evaluation procedures to evaluate the effectiveness of their teachers. Is this simply a way to avoid making difficult decisions regarding teacher performance? Simplistic procedures lacking well-defined performance indicators in conjunction with meaningless feedback given by administration are not producing teacher evaluation data that are linked to student achievement. Therefore, these evaluation instruments are insufficient tools for investigating teacher effectiveness.
It is crucial that teacher performance evaluations have clearly defined performance goals that actually link to student achievement. This literature review has given clarity to the components needed to facilitate the link between teacher performance evaluation and student achievement. These components include a holistic approach that reviews all domains of teaching and uses multiple sources of data as well as quality administrative feedback that leads to teacher differentiation, reflection, and professional growth. Additionally, it is critical for school systems to investigate teacher perception of the performance evaluation instrument and implementation of that instrument to begin the process of restructuring and actually link teacher performance evaluation to student achievement.

**Needed additions to existing research.** In the literature review conducted by Watson, Miller, Davis, and Carter (2012), researchers found a deficiency in studies that have been conducted on the perceptions of teachers and how they view their effectiveness. Watson et al. (2012) used Stronge’s (2007) conceptual framework, which they viewed as the most comprehensive research conducted on teacher effectiveness, as the basis of their study. The study filled a gap in the research by addressing what qualities effective teachers believe good teachers possess. This research provided a more vivid picture of what an effective teacher looks like by adding the teacher’s perspective. However, Watson et al. (2012) stated that additional research needs to be done to unravel the nature of effective teaching. Specifically, Watson et al. (2012) proposed that additional research should focus on whether training programs or teacher evaluation systems actually nurture teaching abilities and create opportunities for teachers to improve and succeed from the perspective of the teachers.
Washington (2011) recommends that school districts conduct their own research to evaluate whether teacher performance predicts student achievement using a correlative design. This would assist school districts in planning and add to the academic knowledge base. Studies assessing evaluation systems and the implications that they have on teacher improvement and classroom instruction would be another valuable addition to the current base of research. Recent research studies investigating the relationship between teacher evaluation and student achievement make it evident that more research is needed to develop a clear picture of the connection.

In order for County X to prepare for the upcoming link between teacher evaluation and student growth data, as well as to prepare for the possibility of pay for performance mandates, it is critical to evaluate whether the current teacher evaluation system is an accurate and reliable tool for measuring teacher effectiveness. This investigation will not only contribute to the existing pool of research attempting to link teacher effectiveness with student achievement, but it will be a first step to accurately align County X teacher evaluation procedures with student achievement data.
Program Description

Chapter 3

Program Overview

County X’s Teacher Evaluation System was created by Carnot et al. (2007) as a method to collect, present, and document data to define performance. County X’s evaluation system is consistent with new requirement recommended by Virginia Department of Education (2011) if adjustments are made to the percentages assigned for sections addressing student growth data to follow the guideline of forty percent. The characteristics contained in the evaluation system place a focus on teacher performance and the academic achievement of their students. County X’s teacher evaluation system collects information from multiple data sources with a goal of producing accountability, professional improvement, and structure to guide common practice while allowing for flexibility and creativity (Carnot et al., 2007).

The primary purposes of the evaluation system are to:

- Improve the quality of instruction by ensuring accountability for classroom performance.
- Contribute to successful achievement of the goals and objectives defined in the vision, mission, and goals of County X.
- Provide a basis for instructional improvement through productive teacher appraisal and professional growth.
- Share responsibility for evaluation between the teacher and the evaluation team in a collaborative process that promotes self-growth, instructional effectiveness, and improvement of overall job performance. (Carnot et al., 2007, p.5)

Carnot, et al. (2007) described a two-tiered approach to defining the responsibilities and major duties of teaching. The expectations for performance are based on job
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

expectations and duties (performance standards) and examples of these observable behaviors (performance indicators). Teachers are evaluated on the following seven performance standards at the formative and summative levels:

- **Data Driven Planning** - The teacher uses data to plan appropriate curricula as well as implement instructional strategies, and uses resources to promote learning for all students.
- **Instructional Delivery** - The teacher promotes learning by addressing individual learning differences and by using effective instructional strategies.
- **Assessment** - The teacher analyzes assessment data to measure student progress and guide immediate and long range instruction.
- **Learning Environment** - The teacher provides a well-managed, safe student-centered environment that is academically challenging.
- **Communication** - The teacher communicates effectively with students, staff, parents/guardians, and the community.
- **Professionalism** - The teacher maintains demeanor, participates in professional growth opportunities, demonstrates an understanding of the curriculum, and contributes to the profession.
- **Student Achievement** – The work of the teacher results in acceptable, measurable student progress. (Carnot, et al. 2007, p.6)

The goal of County X’s Teacher Evaluation System as stated by Carnot et al. (2007) is to collect a comprehensive and fair picture of a teacher’s performance by using the following sources of data: goal setting, observations, documentation log, and students’ survey and summary report. Carnot et al. (2007) uses goal setting as a large component of the teacher evaluation system. Each teacher sets an annual goal that is based on student achievement. The goal is written at the beginning of the school year and monitored at a mid-year review and evaluated for completion at the end of the school year. A goal must be written so that it can be observed and measured. A goal is assessed by its ability to attain SMART status, “SMART stands for specific, measurable, attainable, realistic, and time limited” (Carnot et. al. p. 9). The annual goal is assessed in
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

the student achievement section of the formative and summative evaluations (Carnot, et al. 2007).

Teacher observations conducted by central office instructional staff and building level administration focus on the seven teacher performance standards described above. These observations give information on the teacher’s contributions in the classroom as well as the contributions made to the functioning of the school as a whole. In the evaluation system, the number of observations varies by teacher years of experience and the status of their contracts. First year teachers or teachers who are on an employee improvement plan receive a minimum of four observations and a summative evaluation. Probationary teachers, those teachers who have not achieved continuing contract status, receive a minimum of three observations and a summative evaluation. Continuing contract teachers are placed on a rotation basis in a three year cycle, where every three years they are observed three times and receive a summative evaluation. Continuing contract teachers not in the third year of the rotation do not receive a summative evaluation but have a least one observation (Carnot, et. al 2007).

The third data source used to provide a picture of a teacher’s performance is the documentation log. Carnot et al. (2007) uses the documentation log to allow evaluators to view what they would not normally view with direct observation. The documentation log is continually maintained by the teacher and is brought to all observation conferences as well as the end of the year annual goal meeting. The contents of a documentation log include, but are not limited to, the following components: grading procedures, classroom rules, student survey summary, parent contact log, course syllabus, log of professional development, goal setting form, and all student progress documentation toward meeting
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

the annual goal (Carnot et al. 2007). The documentation log was designed by Carnot et al. (2007) to be a vehicle for self-reflection as teachers document their progress throughout the school year.

The fourth data collection method used by County X teacher evaluation system includes both student surveys and the student survey summary. Carnot et al. (2007) stated that the goal of the student survey is to provide teachers critical information from their students so as to foster continued reflection and improvement. The survey is given by the teacher in the first half of the second nine weeks. Students complete surveys anonymously and return the surveys to the teacher. The survey questions concentrate on student learning, teacher communication and assistance, teacher expectations, classroom environment, and student perception of fairness within the classroom. The teacher maintains the only access to the survey results. Survey data is not verified by administration and is only used for teacher reflection. The teacher reflects on results and summarizes findings in a one page document (student survey summary) that is placed in the documentation log for administrative review (Carnot et al. 2007). The student survey summary requires the teacher to report how many surveys distributed, amount of surveys returned, percentage of questionnaires received, and questions analyzing student satisfaction. (See Appendix A for student survey and student summary survey summary by Carnot et al. 2007, pages 30, 33)

County X Teacher Evaluation System aligns the data sources described with the seven performance standards. Assessment of the quality of a teacher’s performance only happens at the end of the evaluation cycle during the summative evaluation. For probationary teachers or teachers on a plan of improvement this is a yearly process. For
teachers who are on continuing contract status, this occurs every third year of their evaluation cycle (Carnot et al. 2007). This multiple data collection method is used as a source to assign performance ratings on the end of year summative evaluation. A description of terms used for overall performance ratings in the summative evaluation are presented in table 3.1 (Carnot et al., 2007, p. 17).
Table 3.1 County X’s Performance Rating Scale

<table>
<thead>
<tr>
<th>Rating</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceeds Standard</td>
<td>High-quality performance:</td>
</tr>
<tr>
<td></td>
<td>• Exceeds the requirements contained in the job description as expressed in the evaluation criteria</td>
</tr>
<tr>
<td></td>
<td>• Continually seeks opportunities to learn and apply new skills</td>
</tr>
<tr>
<td></td>
<td>• Consistently exhibits behaviors that have a strong positive impact on students and the school climate</td>
</tr>
<tr>
<td></td>
<td>• Serves as a role model for others</td>
</tr>
<tr>
<td>Meets Standard</td>
<td>High–quality performance:</td>
</tr>
<tr>
<td></td>
<td>• Meets the requirements contained in the job description as expressed in the evaluation criteria</td>
</tr>
<tr>
<td></td>
<td>• Demonstrates willingness to learn and apply new skills</td>
</tr>
<tr>
<td></td>
<td>• Exhibits behaviors that have a positive impact on students and the school climate</td>
</tr>
<tr>
<td>Needs Improvement</td>
<td>Inconsistent performance:</td>
</tr>
<tr>
<td></td>
<td>• Requires support in meeting the standards</td>
</tr>
<tr>
<td></td>
<td>• Results in less than quality work performance</td>
</tr>
<tr>
<td></td>
<td>• Leads to areas for professional improvement being jointly identified and planned between the teacher and evaluator.</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>Poor-quality performance:</td>
</tr>
<tr>
<td></td>
<td>• Does not meet the requirements contained in the job description as expressed in the evaluation criteria.</td>
</tr>
<tr>
<td></td>
<td>• May result in the employee not being recommended for continued employment.</td>
</tr>
</tbody>
</table>

(Carnot et al. p.17)
County X’s Teacher Evaluation System contains performance indicators for each of the seven performance standards. Performance indicators are provided by Carnot et al., (2007) as an example of observable behavior that occurs when the standard is met. These sample performance indicators are not an exhaustive list but are provided to teachers and administration to establish clarity and understanding of the performance standard. County X’s Teacher Evaluation System also contains a performance rubric for each performance standard that outlines observable teacher behavior for each of the four performance ratings: exceeds standard, meets standard, needs improvement, and unsatisfactory (see Appendix B for performance indicators and rubrics by Carnot et al., 2007, p. 70-76). Carnot et al. (2007) provided this tool as a resource to administration to guide evaluators in teacher assessment of standard performance.

Using County X’s Teacher Evaluation System, a teacher is placed on a plan of improvement if he/she received two areas ranked as needs improvement or one area ranked as unsatisfactory. Written notice must be given to the teacher of the need for a plan of improvement prior to the summative evaluation. Written notification must also be given for the areas of concern that need to be addressed as well as an employee improvement plan to address these concerns (Carnot et al., 2007).
County X evaluation system is explained using the logic model listed below Table 3.2 (Carnot, et al., 2007):

**Table 3.2 Logic Model of County X Evaluation System**

<table>
<thead>
<tr>
<th>Resources</th>
<th>Activities</th>
<th>Outputs</th>
<th>Short and Long Term Outcomes</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>Formative evaluation</td>
<td>Number of formative evaluations</td>
<td>Teachers will perform effectively within the following performance standards:</td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td>Summative evaluation</td>
<td>Number of summative evaluations</td>
<td>• Data driven planning</td>
<td></td>
</tr>
<tr>
<td>Evaluators</td>
<td>Feedback during summative and formative evaluation conferences</td>
<td>Portfolio collections</td>
<td>• Instructional Delivery</td>
<td></td>
</tr>
<tr>
<td>County X Teacher Training</td>
<td>Documentation log</td>
<td>Number of students surveyed</td>
<td>• Assessment</td>
<td></td>
</tr>
<tr>
<td>County X Teacher Evaluation Policy</td>
<td>Writing annual goal</td>
<td>Teacher’s collection of student data toward annual goal achievement</td>
<td>• Learning Environment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conducting student surveys</td>
<td>Ratings on summative evaluation</td>
<td>• Communication</td>
<td></td>
</tr>
</tbody>
</table>

**Program Training / Validation**

County X does not annually train evaluators that are responsible for conducting teacher performance evaluations. Evaluators and teachers are given a copy of the Teacher Performance Handbook and the handbook is accessible on County X’s website. Teacher performance evaluation procedures are reviewed with new teachers during new teacher training. There have been no efforts from County X to validate the teacher evaluation process.
County X’s procedures are closely aligned with procedures adopted by the state of Virginia. James H. Stronge, Ph.D., The College of William and Mary, acted as project consultant for County X’s development of the Teacher Performance Evaluation Handbook (Carnot et al., 2007). Dr. Stronge also served as a consultant to develop the Virginia Department of Education (2011) guidelines for teacher evaluation. Therefore, training provided by the Virginia Department of Education (VDOE) would be useful and relevant for evaluators in County X. County X does not require evaluators to attend VDOE training.

**Program Evaluation Environment**

Due to Virginia’s recent mandate that 40 percent of teacher evaluation be based on student achievement and the struggle with the state initiated changes to the History SOL test in 2011, Math SOL tests in 2012, and English, Science, and Writing tests in 2013, County X wants the current evaluation system to be assessed. SOL test scores in history and mathematics have dropped significantly since the changes have been established, and the requirements for graduation have not changed: students are still required to pass 6 SOLs for a standard diploma and 9 SOLs for an advanced diploma. With County X’s continual drop in scores, teachers are pressured to provide instruction that is effective on a daily basis. School administration is held accountable for teacher effectiveness and student achievement. This is why it is critical for County X to have a performance evaluation system that is able to accurately assess the effectiveness of teachers. County X must guarantee that they are able to predict teacher effectiveness to ensure that their students are receiving the best possible instruction. This starts with adopting an effective tool to evaluate teachers. Within County X, building administrators
and central office staff are looking for guidance for assessing teachers with these new mandates in place. The questions that this evaluation answers could provide them with valuable data to steer them in the correct direction. This program evaluation has a two-fold purpose to provide an understanding of the impact that the current evaluation system has on teacher practice and to determine whether the current evaluation instrument influences student achievement.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Methodology

Chapter 4

Purpose of the Research

The purpose of this study was two-fold: it investigated County X’s evaluation system to find if the evaluation system could accurately predict student achievement, and it provided an understanding of the impact that the current evaluation system has on teacher practice. First, the determination of whether County X’s teacher evaluation system could accurately predict student achievement was obtained by determining if a strong positive correlation existed between teachers’ summative performance evaluation scores and student performance data. This study also examined the link between administrative holistic teacher performance assessment and the student performance data, as well as the correlation between the administrative holistic performance assessment and teachers’ summative evaluation performance scores. A second focus of this research investigation was to examine the impact that County X’s system had on teacher practice.

This research study was a variation on past research in three critical ways. First, the research study included all aspects of teacher summative evaluation performance. Second, it included a holistic performance assessment based on administrative opinion. Third, it assessed whether the evaluation system influences teacher practice using teacher perspective and opinion. This investigation was the first step in a process to align County X’s evaluation system with effective measurements of teacher performance that accurately predicts student achievement.
Research Questions

Part 1: Performance Evaluation Systems Relationship between Teacher Performance and Student Achievement – In-depth Study of a Subset of Teachers in County X.

1. Do the scores on the teacher summative performance evaluation correlate to the student performance score?
2. Do administrative holistic performance assessments correlate to the student performance score?
3. Do the teacher summative evaluation performance scores correlate to the administrative holistic performance assessment score?

Part 2: The Impact of Performance Evaluations on Teacher Instructional Practice – Survey of High School Teachers in County X.

4. Are teachers who receive meaningful feedback more likely to use the results of performance evaluation for reflection?
5. Are teachers who use the results of the performance evaluation for reflection more likely to report that the teacher evaluation system has a strong impact on their teaching practices?
6. Are teachers who receive meaningful feedback more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

Research Hypothesis

Part 1: Performance Evaluation Systems Relationship between Teacher Performance and Student Achievement - In-depth Study of a Subset of Teachers in County X.
While many researchers have found links between teacher evaluation and student achievement, these non-experimental design studies have only yielded weak correlational values to represent the relationship (Kane et al., 2010; Kimball et al., 2004; Milanowski, 2004; Washington, 2011). In addition, the only experimental design study uncovered that links student achievement to teacher performance revealed that evaluators were not able to accurately judge or effectively differentiate teaching regardless of training or educational level (Strong et al., 2011). Researchers indicated that evaluators are not able to successfully differentiate between teachers causing there to be little variance between performance evaluation ratings (Stronge et al., 2011; Kane et al., 2010; Jacob & Lefgren, 2008; Washington, 2011). This inability to differentiate between teacher performances is reflected in a lack of subsequent variance in data, which subsequently leads to low correlational values. Therefore, this study hypothesized that the teacher summative performance evaluation score will not be correlated with student performance data.

\textbf{H1- Teacher summative performance evaluation score will not be correlated with student performance score.}

A teacher evaluation system is only as strong as its evaluators. Research indicated that evaluators are not able to successfully differentiate between teachers with the majority of teachers receiving the same (satisfactory) rating (Stronge et al., 2011; Kane et al., 2010; Jacob & Lefgren, 2008; Washington, 2011). One of the major reasons attributed to the problem of lack of differentiation in performance evaluations is the inability or unwillingness of evaluators to give strong negative feedback when necessary (Hinchey, 2010; Marx, 2007). Research showed that stronger correlations between teacher performance evaluations and student achievement are found when a more holistic
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

approach is used that includes all of the teacher performance domains as well as multiple sources of data (Jacob & Lefgren, 2008; Kane et al., 2010; Kimball et al., 2004).

By using a holistic performance assessment that incorporates the complete picture of teacher effectiveness and at the same time does not require administration to give negative feedback, division administrators should be able to remove the barriers so that differentiation can occur. Therefore, the administrative holistic performance assessment will correlate with student performance.

**H₂ – Administrators holistic performance assessment score will be correlated positively with the student performance score.**

One of the major problems attributed to lack of differentiation in teacher performance scores is the inability or unwillingness of evaluators to give strong negative feedback when necessary (Hinchey, 2010; Marx, 2007). Patterson, Grenny, McMillan, and Switzler, authors of *Crucial Conversations* (2012), say that people use many tactics to avoid touchy issues and difficult conversations. This lack of differentiation between teachers’ performance evaluation scores may simply be caused by administration avoiding difficult conversations. The administrative holistic performance assessment score will differ from the teacher summative performance evaluation score in that it will not require feedback and discussion. It is hypothesized that administration will give a more honest appraisal of performance when feedback to the teacher is not required. Therefore, it is not expected that the teacher summative performance score will correlate to the administrative holistic performance assessment.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

**H₃ - Teacher summative performance score will not be correlated with administrative holistic performance assessment score.**

**Part 2: The Impact of Performance Evaluations on Teacher Instructional Practice—Survey of High School Teachers in County X**

Teachers believe that more feedback and follow up is needed from the evaluation process (Rothberg & Fenner, 1991; Kennedy, 2012). Additionally, research shows that teachers consistently seek more feedback from the evaluation process (Rothberg & Fenner, 1991). Within the research, teachers consistently stated that the purpose of evaluation should be for reflection and professional growth (Peterson & Comeaux, 1990; Feeney, 2007). This is a systematic challenge as researchers have shown that quality feedback is often a missing component from the teacher evaluation process (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011). Teachers seek meaningful feedback, value feedback as a necessary component in evaluation, and believe that evaluation should be for reflection and professional growth. Therefore, if teachers receive the meaningful feedback requested on performance evaluations, they will use it for reflection. It is therefore hypothesized that if teachers perceive that they receive meaningful feedback during the performance evaluation they will be more likely to report that they use this feedback for reflection.

**H₄ - Teachers receiving meaningful feedback will be more likely to use the results of performance evaluation for reflection.**

Although it was found in the research studies reviewed that teacher performance evaluation did not facilitate reflection in practice (Peterson & Comeaux, 1990; Henson & Hall, 1993; Rothberg & Fenner, 1991; Kennedy, 2012; Derrington, 2011) research
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

indicates that teachers believe that the primary purpose of performance evaluation should be for their own professional growth (Peterson & Comeaux, 1990). Teachers believe that evaluation should facilitate a teacher’s reflection on his or her classroom practice, which should in turn improve a teacher’s classroom skills and promote their professional growth (Peterson & Comeaux, 1990). Teachers believe that if they reflect on their performance evaluation, the result will positively impact their effectiveness as a teacher.

Unfortunately, because essential pieces such as quality feedback and follow up are found to be missing from evaluation procedures (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011) teachers do not always perceive evaluations to be meaningful and worthy of reflection. If teachers actually reflect upon their performance evaluations it is hypothesized that they will more likely report that the teacher evaluation system has a strong impact on their teaching practices.

**H₅-Teachers using results of the performance evaluation for reflection will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.**

Research indicates that an evaluation instrument would only be effective if teachers actually view the instrument as fair and valid (Peterson & Comeaux, 1990; Kyriakides, Demetriou, & Charalambous, 2006; Tyler, 2005). In order for the instrument to be valid, and therefore effective, it must provide quality feedback and follow-up to the teachers (Rothberg & Fenner, 1991; Kennedy, 2012). However, the literature review showed that quality feedback and follow-up, essential pieces, were consistently missing from evaluation procedures (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011), thus making teachers perceive evaluation to be an ineffective formality. Feedback
is the key factor in the perceptions that teachers have in regard to the impact of the evaluation system on their teaching practice. It is therefore hypothesized that teachers receiving meaningful feedback will be more likely to report that the evaluation system has a strong impact on their teaching practices.

\[ \text{H}_6 \quad \text{- Teachers receiving meaningful feedback will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.} \]

**Research Design**

This study employed a cross-sectional non-experimental research design to examine the above hypotheses. During this program evaluation of County X’s teacher evaluation system, data were collected from Algebra I, Algebra II, and Geometry teacher’s summative evaluations from the 2011-2012 and the 2012-2013 school years making the research a cross-sectional descriptive study. The sample reviewed summative evaluation documents of 25 SOL math teachers. This sample included teachers from the three high schools contained within County X. In addition, online surveys were distributed to 234 teachers who were employed in the three high schools in County X. This survey obtained a broad perspective on the impact that the evaluation system has on teacher practice.
Part 1: Performance Evaluation Systems Relationship between Teacher Performance and Student Achievement – In-depth Study of a Subset of Teachers in County X.

Sample

Purposive sampling was used to select Algebra I, Geometry, and Algebra II teachers from County X who had a summative evaluation within the last two years. A two year cycle was selected to review summative evaluations and SOL test data that were completed using the revised math SOL that became effective during the 2011-2012 testing cycle. Mathematics success is critical for students to obtain needed verified credits for graduation. Students must pass one math SOL to graduate with a standard diploma and two math SOLs to graduate with an advanced diploma from high school. The selection of SOL subjects that would satisfy this requirement is Algebra I, Algebra II and Geometry. Scores in these SOL math subjects have drastically declined in County X since the implementation of the revised SOL. It is critical for County X to be able to assess teacher quality and predict student achievement in mathematics.

Further support for the selection of mathematics courses for this study is that Algebra I and Algebra II are one of the leading predictors for college readiness. Riddle (2010) lists success in Algebra I and Algebra II as keys to college readiness in NASSP: The Principals Difference. By preparing students to successfully complete Algebra I and Algebra II, teachers are preparing them to be college and career ready (Riddle, 2010).
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 4.1

*Teachers’ Summative Evaluations*

<table>
<thead>
<tr>
<th>Summative</th>
<th>HS1</th>
<th>HS2</th>
<th>HS3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-2012</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2012-2013</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>8</td>
<td>6</td>
<td>25</td>
</tr>
</tbody>
</table>

**Measures**

**Variables.** Two independent variables, administrators’ holistic performance assessment score and teachers’ summative evaluation performance score, were correlated to the dependent variable, students’ performance scores on the end-of-year Algebra I, Geometry, and Algebra II Standard of Learning Assessment. The administrative opinion scores and the teachers’ summative evaluation scores are an average of a one to four point rating system where a rating of 1 represents unsatisfactory, 2 represents needs improvement, 3 represents meets standard and 4 represents exceeds standard. The dependent variable is an average of the students’ Algebra I, Geometry or Algebra II Standards of Learning score with a range of 200 to 600 assigned points.

summative evaluation report a rating is assigned to the teacher for each of these seven standards. Evaluators choose from the following rating choices: exceeds standard 4, meets standard 3, needs improvement 2, unsatisfactory 1. The evaluator must also provide a narrative to support the rating choice. This investigation calculated an average of the numerical ratings assigned to each of the seven summative categories to give a summative evaluation performance score.

**Administrators’ holistic performance assessment score.** Through interview with the researcher, administration rated the mathematics teachers in their building using a basic 4 point system. Administration verbally reported a rating given the following choices of ratings: exceeds standard (4), meets standard (3), needs improvement (2), and unsatisfactory (1). Administrators only rated teachers whom they had directly observed in the classroom setting and submitted data to their summative performance assessment. The principal and assistant principals from each school participating contributed to this data collection. An average of the scores was taken to obtain the administrators holistic performance assessment score.

**Student performance score.** Student performance score was calculated using End-of-Course Algebra I, Geometry, or Algebra II Standards of Learning (SOL) assessment data. Data were chosen that directly corresponded to the year the summative evaluation was performed. For example, if the teacher’s summative evaluation was conducted in 2011, then the Algebra I, Geometry or Algebra II SOL data were chosen from 2011. This provided a match from the summative evaluation performance score and the student performance score. A students’ performance score was calculated for each teacher by adding up all student SOL scores per subject area (Algebra I, Geometry,
Algebra II) and dividing by the total number of scores. This average of the SOL scores was reported as the student performance score. Data were also collected to report the teacher’s percent of student passing and the teacher’s percent of students passing advanced.

**Summary of measures.** Researchers indicated that evaluators were not able to differentiate between teachers based on their findings of low variance in performance ratings (Stronge et al., 2011; Kane et al., 2010; Jacob & Lefgren, 2008; Washington, 2011). Additionally, research conducted to link teacher evaluation and student performance yielded weak correlational relationships (Kane et al., 2010; Kimball et al., 2004; Milanowski, 2004; Washington, 2011). This study first investigated whether County X’s teacher evaluation data were consistent with the current research. Specifically, the study investigated the correlational relationship between a teacher’s summative performance score and student performance score. The study also examined the variation among summative performance scores.

Researchers agree that evaluations have a lack of variation with most teachers receiving a satisfactory rating (Jacob & Lefgren, 2008; Kane et al., 2010; Stronge, et al., 2011; Washington, 2011). One reason that researchers offer as explanation is the unwillingness of evaluators to give negative feedback (Hinchey, 2010; Marx, 2007). This research investigation explored whether a holistic assessment score that does not require feedback, documentation, or discussion was correlated to the student performance data. The research also investigated whether a holistic assessment score has a higher correlation with student achievement then the summative performance score.
Finally, this study explored the correlation between the summative performance score and the administrators’ holistic performance assessment score to evaluate how closely the two scores are related.

**Analytical Plan**

Correlation coefficients were calculated to compare the summative evaluation score with the administrative holistic performance assessment score, the summative evaluation score with the students’ performance scores, and the administrative holistic performance assessment score and the students’ performance scores. Descriptive statistics including the range, variance, and standard deviation were calculated for the administrative holistic performance assessment scores, student performance scores, and teacher’s summative performance evaluation scores.

**Part 2: The Impact of Performance Evaluations on Teacher Instructional Practice-Survey of High School Teachers in County X**

**Sample**

Surveys were administered to teachers in the three high schools in County X: High School 1 (HS1) 86 teachers, High School 2 (HS2) 72 teachers, and High School 3 (HS3) 76 teachers. Teachers willing to participate chose to complete the survey anonymously and return it to a labeled box in the main office at each high school. The survey instrument was given to teachers at the end of the year faculty meeting after teachers had received their summative evaluation rating for the 2012-2013 school year.

Participants were selected only from the secondary setting because high schools have consistent course requirements and expectations for graduation that are unique to
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

the secondary setting. For example, to graduate high school with a standard diploma a student must complete 24 credits and obtain 6 verified credits by passing 2 English, 1 math, 1 science, and 1 history SOL, as well as 1 SOL of their choosing. Students’ graduation and successful completion of the SOL requirements are governed by a teacher’s classroom achievement.

Instrument

An anonymous survey collected data from one open-ended and 32 close-ended response questions. Close-ended items contained a Likert response scale from one to five. The survey began by collecting demographic information on total years teaching experience, teaching experience in County X, teaching subject, age, gender, and educational level. The survey required the teacher to reflect upon their last summative performance evaluation and answer questions in regard to validity, evaluation feedback, multiple sources of data used, and reflection on practice to improve teaching practice. Additionally, the survey asked teachers to reflect on their most recent summative evaluation experience in County X and rate the overall quality and impact of the evaluation. Finally the survey instrument allowed teachers to write any additional comments they may have about County X’s teacher performance evaluation system as an open ended response question. Survey questions 15-19, 31, 32 were adapted from the Stiggins and Duke (1990) Teacher Evaluation Profile. The survey was administered at the end of the evaluation cycle after the summative evaluations had been conducted. A complete survey is included in Appendix C.
Measures

Researchers stated that evaluation instruments are effective only if they are viewed as fair and valid by the teachers they assess (Peterson & Comeaux, 1990; Kyriakides et al., 2006; Tyler, 2005). Teacher opinion therefore becomes a critical component in the success of an evaluation instrument. Additionally, researchers reported that teachers believe that quality feedback is the key to effective evaluation and professional growth (Rothberg & Fenner, 1991; Kennedy, 2012). Peterson and Comeaux (1990) reported that teachers believe that reflection on practice and professional growth are the main purposes for teacher evaluation. Therefore it was critical to assess whether County X’s evaluation system provided meaningful feedback to the teachers it appraised.

To gauge the impact of County X’s evaluation system, a survey was given to assess feedback and its direct impact on teacher practice. The survey consisted of 32 closed-ended and one open-ended questionnaire item. Closed-ended questions gave participants a rating scale response choice. The survey was given at the end of the evaluation cycle after the summative evaluations had been conducted. These components provide critical information necessary to assess teacher’s performance evaluations impact on teacher instructional practice as outlined in hypotheses.

Analytical Plan

Survey results from closed-ended questions were tallied and the average obtained to produce results from questionnaire data. Responses to survey question 33 were paraphrased into short statements that summarized meaning, statements were reviewed,
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

and common themes to responses were determined. Common themes were developed because of repetitive responses.

Survey data were analyzed using t-test, bivariate correlation and multivariate analysis to investigate items contributing to teachers’ perception of the overall quality of the evaluation and the overall impact of the evaluation on their teaching practice.

IRB

This study fell under the category of exempt review. It involved survey procedures (of adults) and the collection of existing documents and records. The survey results were held confidential and information documented from existing records were recorded so that subjects could not be identified.

Conclusion

This research project was chosen to determine if a problem existed with the current procedure to assess teacher effectiveness. If results revealed that County X evaluation system is not an effective process to influence student performance and modify teacher practice then the next step would be to start refurbishment of the system. A starting place of restructuring if needed would be with County X’s vision statement. Milanowski (2011) defined the starting point as transferring a school system’s vision into standards that explicitly define competent performance. These standards would include rating scales that would clearly describe each level of competence, procedures for clearly evaluating evidence, and training for observers to obtain inter-rater reliability (Milanowski, 2011). Therefore, the next step in the process to align County X’s evaluation system with effective measurement of teacher performance and the ability to
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

influence student achievement would be to determine whether the evaluation system is a reflection of the school system’s vision.
Results of the Study

Chapter 5

Introduction

This chapter contains a presentation and discussion of the results of this research study. A primary purpose of this research study was to investigate County X’s evaluation system to find if the evaluation system influences student achievement and affects teachers’ instructional practices. Therefore, this presentation and discussion will be broken down into two sections, Part 1: Performance evaluation systems relationship between teacher performance and student achievement – in depth study of a subset of teachers in County X, and Part 2: The impact of performance evaluations on teacher instructional practice- survey of high school teachers in County X. Each section in this chapter begins with a description of the research sample and then follows with a discussion of the analysis of data and an examination of the research questions. Part 2: The impact of performance evaluation on teacher instructional practice will also include a multivariate analysis of the survey questions.

The following six research questions guided the study’s analysis.

Part 1: Performance evaluation systems relationship between teacher performance and student achievement – in depth study of a subset of teachers in County X.

1. Do the scores on the teacher summative performance evaluation correlate to the student performance score?

2. Do administrative holistic performance assessments correlate to the student performance score?
3. Do the teacher summative evaluation performance scores correlate to the administrative holistic performance assessment score?

Part 2: The impact of performance evaluations on teacher instructional practice- survey of high school teachers in County X.

4. Are the teachers who receive meaningful feedback more likely to use the results of performance evaluation for reflection?

5. Are teachers who use the results of the performance evaluation for reflection more likely to report that the teacher evaluation system has a strong impact on their teaching practices?

6. Are teachers who receive meaningful feedback more likely to report that the teacher evaluation system has a strong impact on their teaching practices?

Part 1: Performance Evaluation Systems Relationship between Teacher Performance and Student Achievement – In-depth Study of a Subset of Teachers in County X.

This section, which explores a subset of mathematics teachers in County X, will start with a description of the research sample and then follow with a discussion of data and measurement. This section will then provide a discussion of descriptive analysis of data and examine the research questions presented in Part I. Finally, the overall findings from the summative performance score, the administrative holistic performance score, and the student achievement score in County X will be discussed.
Sample

The population of this study included all high school math teachers who taught Algebra I, Geometry, or Algebra II in County X who received a summative evaluation\(^2\) within the last two school years (2011-2012 or 2012-2013). County X employed thirty-two mathematics teachers in three high schools. In this study twenty-five teachers met the aforementioned selection criteria: they taught Algebra I, Geometry, or Algebra II and their administrative staff completed a summative evaluation of their performance within the two year cycle. Therefore, this study includes data from the summative evaluations of twenty-five teachers across County X’s three high schools. Additionally, this study collected data from seven administrators, both principals and assistant principals serving three high schools to obtain the administrator’s holistic performance assessment score (see Appendix D for holistic performance assessment data) on the 25 teachers (N=25). Administrators rated through interview with the researcher the mathematics teachers in their building using a basic 4 point system. Administration verbally reported a rating given the following choices of ratings: exceeds standard (4), meets standard (3), needs improvement (2) and unsatisfactory (1). Administrators only rated teachers whom they had directly observed in the classroom setting and submitted data to their summative performance assessment. Four of the twenty-five teachers selected for this study taught

\[^2\text{In the evaluation system the number of observations varies by teacher years of experience and the status of their contracts. First year teachers or teachers who are on an employee improvement plans receive a minimum of four observations and a summative evaluation. Probationary teachers, those teachers who have not achieved continuing contract status, receive a minimum of three observations and a summative evaluation. Continuing contract teachers are placed on a rotation basis in a three year cycle where every three years they are observed three times and receive a summative evaluation. Continuing contract teachers not in the third year of the rotation do not receive a summative evaluation but have a least one observation (Carnot, et. al 2007).}\]
multiple SOL subjects producing two separate student performance scores (see Appendix E for student performance data). Therefore, the sample includes 29 observations of student performance, average SOL score (N = 29).

**Data and Measurement**

The researcher in this study checked data for accuracy and all responses were within range. The summative performance evaluation score was an additive index comprised of the seven performance dimensions measured in the summative performance evaluation, including data driven planning, instructional delivery, assessment, learning environment, communication, professionalism, and student achievement. On the summative evaluation report a rating is assigned to the teacher for each of the seven standards listed above. Evaluators choose from the following rating choices: exceeds standard (4), meets standard (3), needs improvement (2), unsatisfactory (1). The summative performance evaluation score is an average of the numerical ratings assigned to each of the seven summative categories. The reliability coefficient supports combining the seven criteria to represent the overall construct of performance (Cronbach alpha = .78, mean = 3.03, min= 1.57, max= 3.43). Johnson and Christensen (2012) verify that to show reliability the coefficient alpha for research purposes should be greater than .70.

As noted above, the study asked an administrator to rate teachers whom they had directly observed in the classroom setting on a scale from 1 to 4 (exceeds standard 4, meets standard 3, needs improvement 2, unsatisfactory 1). The administrator verbally reported the rating to the interviewer. For teachers who were observed by more than one
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

The holistic performance rating reported in this study represents an average across the administrators.

The student performance score is an average of the Standard of Learning (SOL) assessment data collected for each subject by teacher. SOL scores range from zero to 600.

**Analysis of Data**

Table 5.1 presents descriptive data for the summative performance evaluation scores, holistic performance assessment score, and the student performance score. As shown in Table 5.1, the mean of the holistic performance score was 2.86, which was lower than the mean of the summative performance score 3.06. Greater variance existed in the holistic performance score than was exhibited in the summative performance score (std = .59, v = .35, std = .36, v = .13, respectively). Administrators differentiated their assessment of performance more using the holistic performance assessment of teachers than they did when using the summative evaluation instrument.

As shown in Table 5.1 the mean student performance score was 398.01 which is below 400, the SOL test pass score. The student performance score had a standard deviation of 43.14, a range of 190.62, a minimum of 307, and a maximum of 497.62. There is a large range for the student performance score from a score of 307, suggesting that no students passed the math SOL test, to a score of 497.62, suggesting that all students passed the math SOL and that most students passed advanced (see Appendix E for student performance data). Given the range of differentiation in the student performance scores across teachers one would expect to see similar differentiation in
teacher performance ratings, especially since research demonstrates that one teacher’s performance can be compared to another teacher in a research environment to determine which teacher has a greater impact on student achievement (Briggs et al., 2012; Borman & Kimball, 2005; Heneman & Milankowski, 2011; Kane, Taylor, Tyler & Wooten, 2010; Mangiante, 2011). However, this was not the case.

The results indicated that 76.0% of the teachers received a meets (3) or exceeds (4) the standard rating on their summative performance evaluation. Table 5.2 presents frequency data for the summative performance evaluation score. In comparison, for the holistic performance score, 64.0% of the teachers received a 3 or higher rating (see Table 5.3). Consistent with the mean analysis above, teachers were rated more highly on the summative evaluations. Specifically, 12% more teachers on the summative evaluations received ratings of meets or exceeds the standard ratings (76%) than teachers did on the holistic performance score (64%). Furthermore, the frequency distributions showed that the majority (76.0%) of the teachers received meets the standard or exceeds standard performance rating although the mean of the student performance score (398.01) was below passing (400).
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 5.1 Descriptive Statistics for Summative Evaluation, Holistic Performance, and Student Performance

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summative Evaluation Performance Score (n=25)</td>
<td>3.06</td>
<td>.38</td>
<td>.14</td>
</tr>
<tr>
<td>Holistic Performance Score (n=25)</td>
<td>2.86</td>
<td>.63</td>
<td>.40</td>
</tr>
<tr>
<td>Student Performance Score (n=29)</td>
<td>398.01</td>
<td>43.14</td>
<td>1861.32</td>
</tr>
</tbody>
</table>

Note: Scores of 1-4 represented performance evaluation descriptions for summative evaluation score and holistic performance score where 4 = Exceeds Standard, 3 = Meets Standard, 2 = Needs Improvement, 1 = Unsatisfactory. The summative evaluation score is an average of the 7 performance standards represented on the summative evaluation instrument. The holistic performance score is a single rating by administration of teaching performance. The student performance score is an average of student SOL scores per SOL subject. SOL scores range from zero to 600, where 400 is passing.

Table 5.2 Frequency of Summative Evaluation Score

<table>
<thead>
<tr>
<th>Summative Evaluation Score</th>
<th>Frequency</th>
<th>Valid Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.57</td>
<td>1</td>
<td>3.8</td>
</tr>
<tr>
<td>2.71</td>
<td>1</td>
<td>3.8</td>
</tr>
<tr>
<td>2.86</td>
<td>4</td>
<td>15.4</td>
</tr>
<tr>
<td>3.00</td>
<td>7</td>
<td>26.9</td>
</tr>
<tr>
<td>3.14</td>
<td>1</td>
<td>3.8</td>
</tr>
<tr>
<td>3.29</td>
<td>7</td>
<td>26.9</td>
</tr>
<tr>
<td>3.43</td>
<td>4</td>
<td>15.4</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note: Scores of 1-4 represent performance evaluation descriptions where 4 = Exceeds Standard, 3 = Meets Standard, 2 = Needs Improvement, 1 = Unsatisfactory. The summative evaluation score is an average of the 7 performance standards represented on the summative evaluation instrument.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 5.3 Frequency of Holistic Performance Score

<table>
<thead>
<tr>
<th>Holistic Performance Score</th>
<th>Frequency</th>
<th>Valid Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.330</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>2.000</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>2.125</td>
<td>2</td>
<td>8.0</td>
</tr>
<tr>
<td>2.330</td>
<td>2</td>
<td>8.0</td>
</tr>
<tr>
<td>2.375</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>2.500</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>2.875</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>3.000</td>
<td>11</td>
<td>44.0</td>
</tr>
<tr>
<td>3.125</td>
<td>1</td>
<td>4.0</td>
</tr>
<tr>
<td>3.750</td>
<td>2</td>
<td>8.0</td>
</tr>
<tr>
<td>4.000</td>
<td>2</td>
<td>8.0</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note: Scores of 1- 4 represent performance evaluation descriptions where 4 = Exceeds Standard, 3 = Meets Standard, 2= Needs Improvement, 1= Unsatisfactory. The holistic performance score is an average of all administrative assessments. The complete set of data including all scores used to calculate the holistic performance score is given in Appendix D.

Table 5.4 presents a descriptive analysis of each of the seven teacher performance standards. Each of the twenty five (N=25) teachers received scores for 7 performance categories, which is a total of 175 ratings. The data revealed that there were only five “unsatisfactory” scores and only fifteen “needs improvement” ratings out of the 175 ratings collected. The largest percentage of teachers receiving needs improvement or unsatisfactory was in the category of student achievement with 36% of teachers receiving a rating that was less than meeting the standard. All assessment standards except student achievement had 93%+ of teachers receiving a 3 “meets standard” or 4 “exceeds standard”. In the student achievement category only 64% of teachers received a 3 or above. The highest percentage of “exceeds standard” was awarded to the category of professionalism with 40% of teachers receiving an “exceeds
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

standard” score. The second largest “exceeds standard” category rating was 32% under instructional delivery.

In summary, the holistic performance score had a lower mean, a larger standard deviation, and a larger range than the summative teacher performance score (mean = 2.86, sd = .59, range= 2.60, mean = 3.06, sd =.36, range = 1.86, respectively). These results suggest that principals differentiated more on the holistic performance assessment than they did when assessing performance using the summative evaluation instrument. Although more differentiation occurred on the holistic performance assessment, it was not to the degree expected considering the wide variation found in the student performance scores (mean 398.01, sd = 43.1, range = 190.6). The wide range on the student performance score and the mean of the student performance score being below passing (400) does not align with the fact that 76% of the teachers received a meets or exceeds standard on their summative performance evaluation. Teacher performance has a direct effect on student achievement (Briggs et al., 2012; Kane et al., 2010; Brophy, 1986). If the mean of the student performance score was below passing, one would expect the mean of the teacher performance evaluation results to be below satisfactory (meets standard), which was not found in this study. Evaluation systems that make connections between classroom instruction and student learning gains will more effectively align teaching with student achievement (Brophy, 1986).
Table 5.4 Percent and Number of Teachers by Performance Standard

<table>
<thead>
<tr>
<th>Performance Standard</th>
<th>Mean</th>
<th>Exceeds Standard</th>
<th>Meets Standard</th>
<th>Needs Improvement</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard 1: Data Driven Planning (n=25)</td>
<td>2.96</td>
<td>4.0%</td>
<td>88.0%</td>
<td>8.0%</td>
<td>0%</td>
</tr>
<tr>
<td>Standard 2: Instructional Delivery (n = 25)</td>
<td>3.24</td>
<td>32.0</td>
<td>64.0</td>
<td>0</td>
<td>4.0</td>
</tr>
<tr>
<td>Standard 3: Assessment (n =25)</td>
<td>3.00</td>
<td>4.0</td>
<td>92.0</td>
<td>4.0</td>
<td>0</td>
</tr>
<tr>
<td>Standard 4: Learning Environment (n = 25)</td>
<td>3.08</td>
<td>20.0</td>
<td>72.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Standard 5: Communication (n =25)</td>
<td>3.04</td>
<td>12.0</td>
<td>84.0</td>
<td>0</td>
<td>4.0</td>
</tr>
<tr>
<td>Standard 6: Professionalism (n = 25)</td>
<td>3.32</td>
<td>40.0</td>
<td>56.0</td>
<td>0</td>
<td>4.0</td>
</tr>
<tr>
<td>Standard 7: Student Achievement (n = 25)</td>
<td>2.68</td>
<td>8.0</td>
<td>56.0</td>
<td>32.0</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Examination of the Research Questions

The purpose of this study was to investigate County X’s evaluation system to find if the teacher evaluation system influences student achievement. In order to examine if the evaluation system was able to effectively measure teacher performance and influence student achievement, this study explored relationships between teachers’ summative performance evaluations, holistic teacher performance assessment, and student performance data.

The following research questions framed this study:

1. Do the scores on the teacher summative performance evaluation correlate to the student performance score?
H₁ - Teacher summative performance evaluation score will not be correlated with student performance score.

2. Do administrative holistic performance assessments correlate to the student performance score?

H₂ – Administrators holistic performance assessment score will be correlated positively with the student performance score.

3. Do the teacher summative evaluation performance scores correlate to the administrative holistic performance assessment score?

H₃ - Teacher summative performance score will not be correlated with administrative holistic performance assessment score.

This study used correlational analysis to address research questions 1-3 listed above. Table 5.5 shows the bivariate Pearson correlations for teacher summative performance evaluation score, administrative holistic performance assessment score, and student performance scores.
Table 5.5 Pearson correlations between teacher summative performance score, administrative holistic performance assessment score, and student performance score.

<table>
<thead>
<tr>
<th>Administrative Holistic Performance Score</th>
<th>Summative Performance Score</th>
<th>Student Performance Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pearson Correlation</strong></td>
<td><strong>Significance</strong></td>
<td></td>
</tr>
<tr>
<td>Administrative Holistic Performance Score</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Summative Performance Score</td>
<td>.760 **</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N = 25)</td>
<td></td>
</tr>
<tr>
<td>Student Performance Score</td>
<td>.265</td>
<td>.395*</td>
</tr>
<tr>
<td></td>
<td>.165</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>(N = 29)</td>
<td>(N = 29)</td>
</tr>
</tbody>
</table>

Note: Data were collected from the summative evaluations of twenty-five teachers to obtain results of the summative performance evaluation score (N=25). Four of the twenty-five teachers selected for this study taught multiple SOL subjects producing two separate student performance scores. Therefore, the student performance score resulted in twenty-nine (N=29) total observations.

*Correlation is significant at the 0.01 level when a two tailed test is used and
**Correlation is significant at the 0.05 level when a two tailed test is used.

Research question number 1 asked: Do the scores on the teacher summative performance evaluation correlate to the student performance score? The correlation between summative performance evaluation and student performance score was found
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

to be positive, significant, and moderate $r (29) = .395, p = .034$. These results did not support the hypothesis that teacher summative evaluation scores will not be correlated with the student performance score.

The findings for research question 1 are inconsistent with previous research studies investigating the link between teacher performance evaluations and student achievement. Previous research studies have only yielded weak correlational values to represent the relationship between teacher performance evaluations and student achievement (Kane et al., 2010; Kimball et al., 2004; Milanowski, 2004; Washington, 2011). This study found a positive, significant, and moderate correlation between teacher performance evaluations and student achievement $r = .395, p = .034$. The significant correlation between teacher summative performance evaluation and the student performance score validates County X’s teacher evaluation system. Goe (2013) stated that it is important for principals to have data to support their conclusions on performance evaluations in order to provide meaningful, valid performance evaluations that are linked to student achievement. The findings of a significant positive correlation between teacher’s summative performance evaluations and student performance data is important to County X administration as well as other school districts who are considering adoption of a performance evaluation system similar to County X’s system.

Although prior research investigations have produced weak correlations, the results of these studies have been consistent. The results indicated evaluation systems that use a standards based approach that include multiple sources of data, and produce a positive relationship between teacher performance and student achievement (Hinchey, 2010; Kane et al., 2010; Kimball et al., 2004; Stonge et al., 2011; Ward, Grant, 2011).
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Since County X’s evaluation system includes a standards based approach that considers multiple sources of data, it makes sense that the results of this investigation support the findings from Hinchey (2010), Kane et al., (2010), Kimball et al., (2004), and Stronge et al., (2011).

Additionally, researchers have indicated that evaluators were not able to successfully differentiate between teachers resulting in evaluation scores that have limited variance, which subsequently leads to low correlational values (Stronge et al., 2011; Kane et al., 2010; Jacob & Lefgren, 2008; Washington, 2011). The findings in this investigation (Table 5.2: Frequency of Summative Evaluation Score) showed that 72.4% of the teachers received a 3 or higher rating (meets standard or exceeds standard). The data collected for summative evaluation performance score showed a mean of 3.05, a standard deviation of .356, and a variance of .126. Although the correlations found in this investigation \( r(29) = .395, p = .034 \) were positive and significant, they were still relatively moderate, likely due to the lack of variance in performance data. Therefore, for County X to improve upon its moderate correlational relationship between teacher evaluation and student performance data it must examine the lack of variance in County X performance data as a possible target for improvement. Variance in this research investigation is truncated since the instrument is scaled but the distribution of values showed a skew towards the positive.

Research question number 2 asked: Do administrative holistic performance assessments correlate to the student performance score? The correlation between administrative holistic performance assessments and the student performance score was positive \( r(29) = .265 \) but not significant \( p = .165 \). These results refuted the hypothesis
that the administrative holistic performance assessment score would be correlated positively with the student performance score.

Jacob & Lefgren (2008) demonstrated stronger correlations between teacher performance evaluations and student achievement when a more holistic approach or overall impression of performance was used. One of the major reasons attributed to the problem of lack of differentiation in standards based performance evaluations has been the inability or unwillingness of evaluators to give strong negative feedback when necessary (Hinchey, 2010; Marx, 2007). Jacob and Lefgren (2008) allowed school administration to provide a holistic rating of teacher effectiveness instead of using standards based performance instrument and produced results that linked teacher performance evaluation to student’s achievement without the problem of lack of variation in teacher performance scores. Therefore, this study hypothesized that the administrative holistic assessment would have a significant, positive correlation with the student performance score because the holistic performance assessment captured an overall picture of teacher effectiveness and did not require any feedback to the teacher from the administrator.

The hypothesis was not supported. This investigation did not find a significant positive correlation $r = .265$ $p= .165$ between administrative holistic performance score and the student performance score. As noted previously, the study did find a positive and significant correlation between teacher summative performance evaluation score and the student performance score $r= .395$, $p= .034$. The results from this study indicate that the relationship between summative performance evaluation and student performance score were stronger than the relationship between holistic performance assessment and the
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

student performance score. The findings from this research investigation signify that the summative performance evaluation procedure in County X was more effective in predicting student performance than administrative opinion alone. This research provides additional support for the use of a standard based approach that uses multiple sources of data as predictor of student performance than a holistic administrative appraisal of a teacher’s performance. The findings for research question 2 are consistent with studies by Hinchey, (2010), Kane et al., (2010), Kimball et al., (2010), and Stronge et al., (2011) who found a standard based approach that includes multiple sources of data, produce a stronger relationship between student achievement and teacher performance. The results from this study, based on math teachers in County X, strengthen the validity of the existing teacher performance evaluation system used by County X.

To further investigate the relationship between the holistic performance assessment results and individual performance standards Table 5.6 shows the correlation between the seven summative performance standards and the holistic performance assessment. Data from Table 5.6 shows that there were strong positive significant correlations between holistic administrative assessment and the following performance standards:

- **Instructional Delivery** \( (r = .700, p = .000) \) - The teacher promotes learning by addressing individual learning differences and by using effective instructional strategies.

- **Data driven planning** \( (r = .477, p = .016) \) – The teacher uses data to plan appropriate curricula, implement instructional strategies, and uses resources to promote learning for all students.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

- **Learning environment** (r = .643, p = .001) - The teacher provides a well-managed, safe student-centered environment that is academically challenging.

- **Professionalism** (r = .758, p = .000) - The teacher maintains demeanor, participates in professional growth opportunities, demonstrates an understanding of the curriculum, and contributes to the profession.

- **Communication** (r = .549, p = .005) - The teacher communicates effectively with students, staff, parents/guardians, and the community.

However, the scores between the holistic performance assessment and both of the standards related to student achievement and assessment were not significant.

- **Assessment** (r = .200, p = .337) – The teacher analyzes assessment data to measure student progress and guide immediate and long range instruction.

- **Student Achievement** (r = .347, p = .089) – The work of the teacher results in acceptable, measurable student progress. (Carnot et al. 2007, p.6)
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 5.6 Paired Sample Correlations (Seven Summative Performance Indicators and Holistic Administrative Performance Score)

<table>
<thead>
<tr>
<th>Pair</th>
<th>Holistic Performance Assessment Score &amp; Overall summative evaluation (N = 25)</th>
<th>Correlation</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>.760</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Pair 2</td>
<td>Holistic Performance Assessment Score &amp; Student Achievement (N = 25)</td>
<td>.345</td>
<td>.089</td>
</tr>
<tr>
<td>Pair 3</td>
<td>Holistic Performance Assessment Score &amp; Instructional Delivery (N = 25)</td>
<td>.700</td>
<td>.000</td>
</tr>
<tr>
<td>Pair 4</td>
<td>Holistic Performance Assessment Score &amp; Data Driven Planning (N = 25)</td>
<td>.477</td>
<td>.016</td>
</tr>
<tr>
<td>Pair 5</td>
<td>Holistic Performance Assessment Score &amp; Assessment (N = 25)</td>
<td>.200</td>
<td>.337</td>
</tr>
<tr>
<td>Pair 6</td>
<td>Holistic Performance Assessment Score &amp; Learning Environment (N = 25)</td>
<td>.643</td>
<td>.001</td>
</tr>
<tr>
<td>Pair 7</td>
<td>Holistic Performance Assessment Score &amp; Professionalism (N = 25)</td>
<td>.758</td>
<td>.000</td>
</tr>
<tr>
<td>Pair 8</td>
<td>Holistic Performance Assessment score &amp; Communication (N = 25)</td>
<td>.549</td>
<td>.005</td>
</tr>
</tbody>
</table>

Note: Administrative holistic performance assessment score was paired with each of the seven performance categories that are included in the summative performance evaluation.

Holistic performance assessment score did not correlate with the summative performance standards of assessment and student achievement. The holistic performance evaluations offer less value to the administration as they are not linked to student performance data or to administrators’ assessments of how teachers perform on criteria related to student assessment and achievement. Therefore, findings from this study suggest that using a standards based approach with multiple sources of data is essential if
schools want an approach that holds teachers accountable for student assessment and achievement.

These findings are also consistent with the Goe (2013) study, which found that principals can no longer claim that they simply know good teaching by seeing it. Principals must have the data to support their conclusions on performance evaluations in order to provide meaningful, valid performance evaluations that are linked to student achievement (Goe, 2013). This investigation confirmed that County X’s summative performance evaluation was a better tool for assessing teacher’s performance based on student achievement than the administrative holistic performance score which was based on opinion alone.

Research question number 3 asked: Do the teacher summative evaluation performance scores correlate to the administrative holistic performance assessment score? From the collection of data on 25 mathematics teachers, results showed that the teacher summative evaluation performance scores and the administrative holistic performance assessment of teacher performance were significantly related, \( r (25) = .76, p < .01 \). The results did not support hypothesis 3, teacher summative performance scores are not correlated with administrative holistic performance scores.

The review of literature did not yield any studies that directly correlated summative performance assessment and a holistic performance assessment. Previous literature did attribute the lack of differentiation in teacher performance scores to the inability or unwillingness of evaluators to give strong negative feedback when necessary (Hinchey, 2010; Marx, 2007; Patterson et al., 2012). Therefore, this study hypothesized
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

that since the administrative holistic performance score did not require direct feedback with teachers, administrators would give a more honest appraisal of a teacher’s performance than the summative performance evaluation score. The findings in this investigation did not support these conclusions. The administrative holistic performance score had a strong positive correlation $r = .76$, $p < .01$ with the summative performance evaluation score. These findings indicated that the lack of differentiation in performance scores in this investigation do not appear to be attributed to the inability of evaluators to give strong negative feedback as indicated by Hinchey (2010) and Marx (2007).

Another possible explanation for the findings in this research investigation is that neither the administrative holistic performance assessment nor the summative performance evaluation score required the administration to give strong negative feedback. County X’s summative evaluation requires administration to give a rating for each of the seven performance standards but does not require an overall rating on the performance assessment. Ratings received on the majority (four out of the seven) performance standards determines the teacher’s overall rating but this rating is not written on the performance document therefore it is not communicated directly to the teacher.

Although this study showed that the summative scores were a better predictor of student achievement, the significant positive correlation between the summative performance scores and the holistic performance assessment scores also shows that the holistic performance evaluation scores hold value in the evaluation process. These informal assessments (holistic performance evaluation) blended with formal assessments (summative performance evaluation) may give a better overall picture of teacher performance. Research supports this conclusion by stating that a combination approach
to measuring teacher performance is superior to a single method approach (Martinez-Rizo, 2012).

However, when examining a paired t-test (results reflected in table 5.7), this study demonstrated that the mean difference between the administrative holistic performance assessment score (mean = 2.86) and the summative performance evaluation (mean = 3.06) score was -.20. This difference was statistically significant (T= -2.366, n=25, p=.03). Therefore, a significant difference did exist between the administrative holistic performance assessment score and the summative performance score. In order to explore these scores further the research investigation conducted a paired t-test for the administrative holistic performance assessment score (mean = 2.86) and the seven categories that comprised the summative performance evaluation score. The following standards: student achievement (mean = 2.68), instructional delivery (3.24), data driven planning (mean = 2.96), assessment (mean =3.00), learning environment (mean = 3.08), professionalism (mean = 3.32) and communication (mean = 3.04) were analyzed to assess if a statistically significant difference existed between the holistic performance assessment score and each of the categories that comprised the summative evaluation assessment. The holistic performance score was significantly lower than the scores for the standards of instructional delivery (T= -3.735, n=25, p=.00) and professionalism (T= -4.915, n=25, p=.00). These results indicated that administrators rated teachers significantly higher on the summative evaluation assessment in the categories of instructional delivery and professionalism than the same administrators rated teachers using the holistic performance assessment.
Table 5.7  Paired sample statistics for administrative holistic performance assessment and summative evaluation assessment

<table>
<thead>
<tr>
<th>Pair</th>
<th>Holistic &amp;</th>
<th>Overall Summative Evaluation</th>
<th>Mean</th>
<th>Correlation</th>
<th>df</th>
<th>t</th>
<th>Sig (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Holistic &amp;</td>
<td>Overall Summative Evaluation</td>
<td>2.86</td>
<td>.760</td>
<td>24</td>
<td>-2.366</td>
<td>.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.06</td>
<td>Sig = .000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Holistic &amp;</td>
<td>Student Achievement</td>
<td>2.86</td>
<td>.347</td>
<td>24</td>
<td>1.220</td>
<td>.234</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.68</td>
<td>Sig = .089</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Holistic &amp;</td>
<td>Instructional Delivery</td>
<td>2.86</td>
<td>.700</td>
<td>24</td>
<td>-3.735</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.24</td>
<td>Sig = .000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Holistic &amp;</td>
<td>Data Driven Planning</td>
<td>2.86</td>
<td>.477</td>
<td>24</td>
<td>-0.856</td>
<td>.400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.96</td>
<td>Sig = .016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Holistic &amp;</td>
<td>Assessment</td>
<td>2.86</td>
<td>.200</td>
<td>24</td>
<td>-1.059</td>
<td>.300</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.00</td>
<td>Sig = .337</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Holistic &amp;</td>
<td>Learning Environment</td>
<td>2.86</td>
<td>.643</td>
<td>24</td>
<td>-2.005</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.08</td>
<td>Sig = .001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Holistic &amp;</td>
<td>Professionalism</td>
<td>2.86</td>
<td>.758</td>
<td>24</td>
<td>-4.915</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.32</td>
<td>Sig = .000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Holistic &amp;</td>
<td>Communication</td>
<td>2.86</td>
<td>.549</td>
<td>24</td>
<td>-1.562</td>
<td>.131</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.04</td>
<td>Sig = .005</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Summary

Overall findings from the summative performance evaluation scores, administrative holistic performance scores, and student achievement scores in County X revealed that the summative evaluation score ($r (29) = .395, p= .034$) was a stronger predictor of student achievement than the holistic performance score ($r (29) = .265, p= .165$). This research investigation provided support that the summative performance evaluation procedure in County X which uses a standards based approach and includes multiple sources of data is a stronger predictor of student achievement than the administrative holistic appraisal of teaching performance.

Although research revealed strong positive significant correlation between the holistic performance assessment score and the summative evaluation score ($r (25) = .76, p< .01$), a deeper look at the data demonstrated that there were significant differences between the two performance ratings. Holistic performance assessment score did not correlate with the summative performance standards of assessment ($r = .200, p = .337$) and student achievement ($r = .347, p=.089$). Descriptive statistics revealed that on the summative performance evaluation the standard of student achievement received the lowest overall rating with 38.9% of teachers receiving a needs improvement or unsatisfactory. The holistic performance evaluation offers less value to the administration as they are not linked to student performance data or to administrators’ assessments of how teachers perform on criteria related to student assessment and achievement.

Furthermore, by examining a paired t-test, this study demonstrated that the mean differences between the administrative holistic performance assessment score and the
summative evaluation score were statistically significant (T = -2.366, n=25, p=.03). The study revealed that mean scores were significantly higher on the summative performance assessment than the holistic performance assessment in the categories of professionalism and instructional delivery. Additionally, descriptive statistics supported this finding by showing that the highest percentage of teachers received a score of exceeds standard in categories of instructional delivery and professionalism.

Although the summative evaluation instrument in County X produced a significant correlation with the student performance score, there is concern that teachers received the highest ratings in the category of instructional delivery and the lowest ratings in student achievement on the summative performance assessment. It is unexpected that the categories of instructional delivery and student achievement would be in opposition to each other. Teacher performance has a direct effect on student achievement (Briggs et al., 2012; Kane et al., 2010; Brophy, 1986). The data also reveals additional concerns with 76% of teachers receiving ratings of meets or exceeds standard. The average performance of teachers (mean = 3.06 which implies meets standard) did not align with the average performance on their math SOLs (mean = 398.01 which is below passing score of 400).

**Part 2: The impact of Performance Evaluations on Teacher Instructional Practice – Survey of High School Teachers in County X**

This performance evaluation research also investigated the impact that County X’s evaluation system has on a teacher’s instructional practice. This investigation was conducted by surveying teachers in the three high schools in County X. This section begins with a description of the research sample which includes survey participants’
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

demographic information and then provides a descriptive analysis of data for each the five remaining sections of the survey instrument which include: validity, evaluation feedback, multiple sources of data, reflection on practice, and overall rating. Following the descriptive analysis this section analyzes the results of the open ended survey response question and then uses difference of means, correlational and multivariate analysis to investigate the overall rating of the teacher evaluation system. This section concludes with a summary of the findings.

Sample

For this study, the targeted survey population totaled 234 high school teachers in County X. Out of the total targeted population, 115 participants responded (49%). The survey was comprised of 32 closed ended and one open ended questions (see Appendix C for complete survey). In the first section of the survey, participants provided demographic information. The seven questions in the demographic section related to the respondents’ teaching experience in County X, total years of teaching experience, current teaching assignment, gender, age, level of education and the score they had received on their last summative evaluation.

Survey questions four and six received 115 responses. These survey questions looked at gender and level of education. Question four categorized respondents between male and female. In question four, 76 respondents, or 66.1%, identified themselves as female, while only 39, or 33.9%, labeled themselves as male. In question six, 55, or 47.8%, of the respondents reported their level of education as a bachelor’s degree. The largest group of participants 58, or 50.4% reported the level of education as a master’s degree and 2 respondents, or 1.7% reported holding a doctorate degree. Survey question
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

five, age, received 114 responses. The largest group of participants (30.7%) were between the ages of 41 and 50. Table 5.8 shows the age of respondents.

Table 5.8 Participant Age by Category

<table>
<thead>
<tr>
<th>Age Ranges</th>
<th>Participant Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage</td>
</tr>
<tr>
<td>21-30</td>
<td>21.1</td>
</tr>
<tr>
<td>31-40</td>
<td>23.7</td>
</tr>
<tr>
<td>41-50</td>
<td>30.7</td>
</tr>
<tr>
<td>51-60</td>
<td>18.4</td>
</tr>
<tr>
<td>61+</td>
<td>6.1</td>
</tr>
</tbody>
</table>

In demographic questions one and two, participants selected from a list of options that best described their number of years of experience as a teacher in County X and their total years of teaching experience. The data displayed in Table 5.9 represents the responses of 115 teachers to question one and two. Additionally, question three had participants select the department that they were most closely associated with in their current teaching assignment. The data displayed in Table 5.10 represents the responses of the 113 respondents to question three, current teaching assignment.
Table 5.9  Participant Teaching Experience

<table>
<thead>
<tr>
<th>Years of Experience by Range</th>
<th>County X</th>
<th>Total Experience Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage</td>
<td>Count Total</td>
</tr>
<tr>
<td>1-3 years</td>
<td>21.7%</td>
<td>25</td>
</tr>
<tr>
<td>4-7 years</td>
<td>21.7</td>
<td>25</td>
</tr>
<tr>
<td>8-11 years</td>
<td>21.7</td>
<td>25</td>
</tr>
<tr>
<td>12-19 years</td>
<td>27.8</td>
<td>32</td>
</tr>
<tr>
<td>20 + years</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 5.10  Current Teaching Assignment

<table>
<thead>
<tr>
<th>Department</th>
<th>Percentage</th>
<th>Total Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>12.4%</td>
<td>14</td>
</tr>
<tr>
<td>Science</td>
<td>16.8</td>
<td>19</td>
</tr>
<tr>
<td>English</td>
<td>14.2</td>
<td>16</td>
</tr>
<tr>
<td>Social Studies</td>
<td>9.7</td>
<td>11</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>8.8</td>
<td>10</td>
</tr>
<tr>
<td>Related Arts</td>
<td>15.0</td>
<td>17</td>
</tr>
<tr>
<td>Health PE</td>
<td>5.3</td>
<td>6</td>
</tr>
<tr>
<td>Special Education</td>
<td>17.7</td>
<td>20</td>
</tr>
</tbody>
</table>

On survey question seven, participants were asked to report the score they received on their last summative evaluation. Of the 115 participants, 112 responded to this question. Findings indicated that 98.2% of teachers reported receiving a four (exceeds standard) or a three (meets standard) rating. Four of the 112 respondents reported that they did not know the score of their summative evaluation. It was expected that more respondents would have chosen the response of “do not know score” since County X does not report an overall summative rating on the evaluation document.
As mentioned above, teachers only received scores for each of the seven performance categories with no overall rating. Teachers must know that the majority or four out of seven determined the overall rating. The data displayed in Table 5.11 represents the responses of the 108 respondents to question seven that indicated a summative performance rating.

Table 5.11  Summative Evaluation Score

<table>
<thead>
<tr>
<th></th>
<th>Percentage</th>
<th>Total Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceeds Standard (4)</td>
<td>30.6%</td>
<td>33</td>
</tr>
<tr>
<td>Meets Standard (3)</td>
<td>67.6%</td>
<td>73</td>
</tr>
<tr>
<td>Needs Improvement (2)</td>
<td>.9%</td>
<td>1</td>
</tr>
<tr>
<td>Unsatisfactory (1)</td>
<td>.9%</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>108</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.8</td>
<td></td>
</tr>
</tbody>
</table>

Analysis of Data

In addition to demographic information collected on the survey there were five additional survey sections which include: section 2 (Validity), section 3 (Evaluation Feedback), section 4 (Multiple Sources of Data Used), section 5 (Reflection on Practice), and section 6 (Overall Rating).

- **Validity** – this section asks the participant to respond to questions judging the accuracy and fairness of the evaluation instrument. It also questions whether time spent and information reviewed is adequate and sufficient to judge performance.

- **Evaluation Feedback** – this section asks the participant to respond to questions about the amount, quality, timing, depth and use of information received during a post observation conference.

- **Multiple Source of Data Used** – this section asks the participant to tell whether administrators evaluating their performance used
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district observation, artifacts in the documentation log, student survey summary, and goal setting to assess student achievement.

- **Reflection on Practice** – this section asks the participant to look back over the last year and report the extent the following sources of performance information caused them to reflect and improve.

- **Overall Rating** – this section ask the participant to rate the overall quality of the evaluation and the overall impact the evaluation had on their teaching practice.

The findings from each of these sections will be discussed below (see Appendix C for complete survey).

**Validity.** Section two of the survey asked respondents to reflect on their last summative performance evaluation in County X and respond to questions relating to the validity of the teacher performance evaluation instrument. Participants responded to seven questions on a 1-5 Likert scale, with 1 representing strongly disagree to 5 representing strongly agree. Table 5.12 provides a list of questions related to validity. The data displayed in Table 5.13 represents the responses to questions related to validity on the survey instrument.

Findings from survey data indicated that the majority of respondents agreed that their summative evaluation accurately portrayed their performance (75.2% agreement), clear examples were given to justify ratings on the summative evaluation (71.3% agreement), and that the ratings did not reflect bias from their evaluator (75.6% agreement). Teachers showed agreement that the score they received on their summative evaluation was accurate and justified.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

However, the majority of respondents did not agree that the number of observations allows for an accurate prediction of performance (49.6% agreement), that the goal setting process provided an accurate picture of the teachers ability to impact student learning (45.2% agreement), and that the information collected in the documentation log was a valid way of providing a comprehensive portrait of their work (49.6%).

The generalized conclusions drawn from the results of the validity section suggested that teachers believed that their summative scores were an accurate portrayal of their performance (98.2% of respondents reporting they received a meets or exceeds standards rating) but the tools used such as observation, goal setting process, and the collection of evidence in the documentation log were not perceived by teachers as instruments that produced an accurate assessment of their teaching performance. In addition, when respondents were asked to report in question 14 whether the evaluation instrument used to rate their performance was fair and valid, 55.7% of respondents reported agreement. The question that was not completely understood is whether the agreement comes from the score itself or the evaluation process. There are conflicting results in the validity section of the questionnaire that indicate that teachers report the process to be fair and valid because of the elevated scores and not because of the validity of the process.
Table 5.12 Questions in Section Two on the Survey Instrument – Validity

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>VQ8</td>
<td>The results of my summative evaluation accurately portrayed my performance.</td>
</tr>
<tr>
<td>VQ9</td>
<td>Clear examples were given to justify the ratings I received on my summative evaluation.</td>
</tr>
<tr>
<td>VQ10</td>
<td>The ratings I received on my summative evaluation did not reflect bias from my evaluator.</td>
</tr>
<tr>
<td>VQ11</td>
<td>The number of observations and times spent conducting classroom observations including pre and post observation conferences allows for an accurate prediction of my teaching performance.</td>
</tr>
<tr>
<td>VQ12</td>
<td>The measurement of student performance that is evaluated during the goal setting process provided an accurate picture of my ability to impact student learning.</td>
</tr>
<tr>
<td>VQ13</td>
<td>The information collected in the documentation log provides evidence of several performance standards. The items are a valid way of providing a comprehensive portrait of my work.</td>
</tr>
<tr>
<td>VQ14</td>
<td>The teacher evaluation instrument used to rate my performance is fair and valid.</td>
</tr>
</tbody>
</table>

Table 5.13 Rating Response to Section Two Validity Questions

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percent Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>VQ8 (N=113)</td>
<td>39.8%</td>
<td>35.4%</td>
<td>12.4%</td>
<td>8.8%</td>
<td>3.5%</td>
<td>3.99</td>
<td>1.10</td>
<td>75.2%</td>
</tr>
<tr>
<td>VQ9 (N=111)</td>
<td>34.2</td>
<td>39.6</td>
<td>16.2</td>
<td>7.2</td>
<td>3.6</td>
<td>3.93</td>
<td>1.05</td>
<td>71.3</td>
</tr>
<tr>
<td>VQ10 (N=111)</td>
<td>49.5</td>
<td>28.8</td>
<td>9.9</td>
<td>7.2</td>
<td>4.5</td>
<td>4.12</td>
<td>1.13</td>
<td>75.6</td>
</tr>
<tr>
<td>VQ11 (N=113)</td>
<td>21.2</td>
<td>29.3</td>
<td>19.5</td>
<td>21.2</td>
<td>8.8</td>
<td>3.33</td>
<td>1.27</td>
<td>49.6</td>
</tr>
<tr>
<td>VQ12 (N=113)</td>
<td>21.2</td>
<td>24.8</td>
<td>30.1</td>
<td>15.0</td>
<td>8.8</td>
<td>3.35</td>
<td>1.09</td>
<td>45.2</td>
</tr>
<tr>
<td>VQ13 (N=112)</td>
<td>17.9</td>
<td>33.0</td>
<td>33.0</td>
<td>9.8</td>
<td>6.3</td>
<td>3.46</td>
<td>1.22</td>
<td>49.6</td>
</tr>
<tr>
<td>VQ14 (N=113)</td>
<td>17.7</td>
<td>38.9</td>
<td>31.9</td>
<td>8.8</td>
<td>2.7</td>
<td>3.60</td>
<td>.97</td>
<td>55.7</td>
</tr>
</tbody>
</table>
Evaluation feedback. Section three of the survey asked respondents to reflect on the feedback they received during their last post observation conference to answer the following questions. Participants responded to eight questions on a 1-5 Likert scale (see table 5.14 for scale designations per question). Table 5.14 provides a list of the questions related to evaluation feedback and the associated response data.
Table 5.14 Rating Response to Section Three Evaluation Feedback

Please reflect on the feedback you received during your last post observation conference to answer the following questions.

<table>
<thead>
<tr>
<th>Amount of information received (Q-15)</th>
<th>None (1)</th>
<th>Small Amount (2)</th>
<th>Average Amount (3)</th>
<th>Above Average Amount (4)</th>
<th>Great Deal (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 114)</td>
<td>2.6%</td>
<td>4.4%</td>
<td>27.2%</td>
<td>45.6%</td>
<td>20.2%</td>
<td>3.76</td>
<td>.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Depth of information Received (Q-16)</th>
<th>Shallow (1)</th>
<th>Low Significance (2)</th>
<th>Significant (3)</th>
<th>High Significance (4)</th>
<th>In-Depth (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 113)</td>
<td>3.5%</td>
<td>8.0%</td>
<td>31.9%</td>
<td>43.4%</td>
<td>13.3%</td>
<td>3.55</td>
<td>.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality of ideas and suggestions contained in the feedback (Q-17)</th>
<th>Low (1)</th>
<th>Low Average (2)</th>
<th>Average (3)</th>
<th>High Average (4)</th>
<th>High (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 114)</td>
<td>4.4%</td>
<td>10.5%</td>
<td>30.7%</td>
<td>37.7%</td>
<td>16.7%</td>
<td>3.52</td>
<td>1.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specificity of information provided (Q-18)</th>
<th>General (1)</th>
<th>Minimal Detail (2)</th>
<th>Average Detail (3)</th>
<th>Significant Detail (4)</th>
<th>Specific (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N=114)</td>
<td>2.6%</td>
<td>9.6%</td>
<td>26.3%</td>
<td>45.6%</td>
<td>15.8%</td>
<td>3.62</td>
<td>.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Timing of the feedback (Q-19)</th>
<th>Delayed (1)</th>
<th>Mostly Delayed (2)</th>
<th>Sometimes Delayed (3)</th>
<th>Rarely Delayed (4)</th>
<th>Immediate (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 113)</td>
<td>3.5%</td>
<td>3.5%</td>
<td>12.4%</td>
<td>43.4%</td>
<td>37.2%</td>
<td>4.07</td>
<td>.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Application of information toward student achievement (Q-20)</th>
<th>Low (1)</th>
<th>Minor (2)</th>
<th>Moderate (3)</th>
<th>Adequate (4)</th>
<th>High (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 114)</td>
<td>5.3%</td>
<td>8.8%</td>
<td>27.2%</td>
<td>39.5%</td>
<td>19.3%</td>
<td>3.59</td>
<td>1.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feedback was useful for my professional development (Q-21)</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 114)</td>
<td>6.1%</td>
<td>13.2%</td>
<td>29.8%</td>
<td>34.2%</td>
<td>16.7%</td>
<td>3.42</td>
<td>1.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feedback was meaningful and assisted me to improve my classroom instruction (Q-22)</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 114)</td>
<td>5.3%</td>
<td>13.2%</td>
<td>29.8%</td>
<td>34.2%</td>
<td>16.7%</td>
<td>3.42</td>
<td>1.10</td>
</tr>
</tbody>
</table>
The majority of respondents indicated (by responding with a rating of 4 or 5) that they received a great deal or above average amount of information on the post observation conference (65%), this information was of high significance or in-depth (56%), and the information was of high average or high quality (54%). Additionally, 71% of teachers responded that the administrators provided information with significant or specific detail, 79% reported that administrators rarely delayed or either provided feedback immediately, and 58% said that the feedback provided on the evaluation was adequate or highly applicable to student achievement. However, only 41% of the teachers responding indicated (by responding with a 4 or 5) agreement that administrator provided feedback that was meaningful and that assisted them to improve their classroom instruction. 50% of teachers reported agreement that feedback was used towards their professional development.

The generalized conclusions drawn from the results of the evaluation feedback section suggest that teachers were receiving a large amount of feedback that was specific and timely; however, the information needed more depth, quality, and relevance towards student achievement for the post observation conference to actually influence a teacher’s professional development or assist them with improvement of their classroom instruction.

**Multiple sources of data used.** Section four of the survey asked respondents to reflect over the last year and report the extent to which the following sources of performance information were considered as part of formative and summative evaluations. A formative evaluation is completed multiple times within the school year with the goal of continued improvement in performance and reflection on practice. A summative evaluation of a teacher’s performance is completed at the end of the year and
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

assembles data from multiple formative evaluations (Carnot et. al. 2007). Participants responded to four questions on a 1-5 Likert scale, with 1 representing not considered to 5 representing used extensively. The data displayed in Table 5.15 represents the responses to questions related to multiple sources of data used. Teachers reported that administration used observation of classroom performance (62% of teachers) and goal setting to assess student achievement (57% of teachers) extensively by responding with a rating of 4 or 5. When asked to rate how the examination of artifacts in the documentation log or the student survey summary was used, only 38% and 34% respectively thought the information was used extensively by administrators. The generalized conclusion drawn from these results is that classroom observations and goal setting processes are the two major components used extensively on the summative evaluation.

Table 5.15 Rating Response to Section Four Multiple Sources of Data Used

<table>
<thead>
<tr>
<th>Question</th>
<th>Not Considered (1)</th>
<th>Rarely Considered (2)</th>
<th>Considered (3)</th>
<th>Frequently Considered (4)</th>
<th>Used Extensively (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation of your classroom performance (Q-23)</td>
<td>0%</td>
<td>8.8%</td>
<td>28.3%</td>
<td>37.2%</td>
<td>25.7%</td>
<td>3.80</td>
<td>.93</td>
</tr>
<tr>
<td>Examination of artifacts in your documentation log (Q-24)</td>
<td>14.4</td>
<td>19.8</td>
<td>26.1</td>
<td>31.5</td>
<td>8.1</td>
<td>2.99</td>
<td>1.19</td>
</tr>
<tr>
<td>Student Survey Summary (Q-25)</td>
<td>24.1</td>
<td>19.6</td>
<td>21.4</td>
<td>28.6</td>
<td>6.3</td>
<td>2.73</td>
<td>1.28</td>
</tr>
<tr>
<td>Goal setting to assess student achievement (Q-26)</td>
<td>3.6</td>
<td>8.0</td>
<td>30.4</td>
<td>38.4</td>
<td>19.6</td>
<td>3.62</td>
<td>1.01</td>
</tr>
</tbody>
</table>

**Reflection on practice.** Section five of the survey asked respondents to look back over the last year and report the extent to which the following caused them to
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district reflect: feedback from evaluator given during post observation conference, collection of artifacts in your documentation log, completion of the student survey and student survey summary, and data analysis to assess student achievement as part of the goal setting process. Participants responded to four questions on a 1-5 Likert scale, with 1 representing little reflection to 5 representing great reflection. The data displayed in Table 5.16 represents the responses to questions related to reflection on practice. Teachers reported (by responding with a rating of 4 or 5) that observation of classroom performance (61%) and student achievement as part of the goal setting process (60%) provided them with the greatest reflection. Teachers indicated that classroom observations and data analysis to assess student achievement as part of the goal setting process were not only the items used most extensively on the summative evaluation but they were also the items that provided them the most reflection to improve teaching practice.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 5.16 Rating Response to Section Five Reflection on Practice

<table>
<thead>
<tr>
<th>Question</th>
<th>Little Reflection (1)</th>
<th>Below Average Reflection (2)</th>
<th>Average Reflection (3)</th>
<th>Above Average Reflection (4)</th>
<th>Great Reflection (5)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q-27) Feedback from evaluator given during post observation conference (N=113)</td>
<td>5.3%</td>
<td>10.6%</td>
<td>21.2%</td>
<td>44.2%</td>
<td>18.6%</td>
<td>3.60</td>
<td>1.07</td>
</tr>
<tr>
<td>(Q-28) Collection of artifacts in your documentation log (N=113)</td>
<td>18.6</td>
<td>20.4</td>
<td>24.8</td>
<td>28.3</td>
<td>8.0</td>
<td>2.87</td>
<td>1.24</td>
</tr>
<tr>
<td>(Q-29) Completion of the student survey and student survey summary (N=113)</td>
<td>19.5</td>
<td>11.5</td>
<td>24.8</td>
<td>30.1</td>
<td>14.2</td>
<td>3.08</td>
<td>1.33</td>
</tr>
<tr>
<td>(Q-30) Data analysis to assess student achievement as part of the goal setting process (N=113)</td>
<td>8.0</td>
<td>8.0</td>
<td>23.0</td>
<td>35.4</td>
<td>25.7</td>
<td>3.63</td>
<td>1.18</td>
</tr>
</tbody>
</table>

**Overall rating.** Section six of the survey asked respondents to reflect on their most recent summative evaluation experience in County X considering the entire evaluation process including observation, goal setting, documentation log, student survey summary, feedback, etc. Participants responded to two questions one on a 1-5 Likert scale, with 1 representing very poor quality to 5 representing very high quality and the other on a 1-5 Likert scale, with 1 representing no impact to 5 representing strong impact. The data displayed in table 5.17 represents the responses to questions related to overall rating on the survey instrument.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 5.17   Overall Rating of the Evaluation Process

<table>
<thead>
<tr>
<th>Question</th>
<th>Very Poor Quality (1)</th>
<th>Poor Quality (2)</th>
<th>Average Quality (3)</th>
<th>High Quality (4)</th>
<th>Very High Quality (5)</th>
<th>Standard Deviation</th>
<th>Mean</th>
<th>Percentage Rating High or Very High Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate the overall quality of the evaluation (n= 112)</td>
<td>2.7%</td>
<td>12.5%</td>
<td>33.9%</td>
<td>39.3%</td>
<td>11.6%</td>
<td>.95</td>
<td>3.45</td>
<td>50.9%</td>
</tr>
<tr>
<td>No Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slight Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above Average Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate the overall impact of the evaluation on your teaching practice (n=112)</td>
<td>7.2%</td>
<td>22.5%</td>
<td>32.4%</td>
<td>28.8%</td>
<td>9.0%</td>
<td>1.08</td>
<td>3.10</td>
<td>37.8%</td>
</tr>
<tr>
<td>No Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slight Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above Average Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strong Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The average rating for the overall quality of the evaluation from the 112 survey respondents was 3.45 indicating that the majority 50.9% reported that the quality of the evaluation was high or very high quality. Additionally, the data revealed that the average rating for the overall impact of the evaluation process on a teacher’s practice was a 3.10, showing the largest selection 32% of teachers reported the evaluation to have an average impact on their teaching practice.

In summary, the validity section of the survey revealed that teachers believed their summative evaluation scores were an accurate portrayal of their performance but the tools used such as observation, goal setting process, and the collection of evidence in the documentation log are not valid ways of assessing their teaching performance. Additionally, teachers indicated that classroom observations and the goal setting process
were the items used most extensively on the summative evaluation and were the parts of the evaluation that provided the most reflection to improve their teaching practice. Teachers also reported that they received a large amount of feedback from administrators that was specific and timely; however, the information provided needed more depth, quality, and relevance towards student achievement for the post observation conference to actually influence a teacher’s professional development or assist them with improvement of their classroom instruction.

The results indicated that feedback must not only be given to teachers; it must be meaningful to improve classroom instruction. The lack of depth, quality, and relevance in feedback that was reported by teachers on the survey instrument may be a contributing factor that impacted the overall rating of survey question 32 (overall impact of the evaluation on teaching practice -38% of teachers rated evaluation as having an above average or strong impact) and survey question 31 (overall quality of the evaluation -51% of teachers rated evaluation as high or very high quality).

It is important to note that high performance ratings (98.2% of teachers reported receiving meets or exceeds standard) may have influenced teacher opinion in the overall quality of the evaluation. Teachers who received good scores on their performance evaluations are more likely to rank the evaluation as high quality. This may be the reason that the overall quality of the evaluation is higher (mean 3.446) than the overall impact of the evaluation process on a teacher’s practice (mean 3.009).

**Common Themes to Open Response Question**

Twenty-eight teachers responded to the open ended survey question out of the 115 teachers who were surveyed. Survey question number 33 asked respondents to “please
write any comments you may have about the County X Teacher Performance Evaluation System below.” Respondents listed a number of issues with County X’s teacher performance evaluation system and these problems fell into the categories of frustration with SOL accountability, insufficient time spent on classroom observations and conferences to accurately predict performance, and the lack of meaning in the evaluation process. Johnson and Christensen (2012) stated that open ended response questions should be coded by examining the survey responses, determining meaningful categories of information, and sorting meaningful responses into inductive categories. Each of the 28 responses to survey question 33 was paraphrased into short statements to summarize meaning, statements were reviewed, and common themes to responses were determined. Three common themes developed because of repetitive responses. These common themes include: frustration with accountability 5 out of 28 respondents (5/28 *100 =18%), insufficient time spend on classroom observations and conferences to accurately predict performance, 9 out of 28 respondents (9/28 *100 = 32%); and lack of meaning in the evaluation process, 7 out of 28 respondents (7/28*100 = 25%). All other responses were not included in the analysis since they only occurred 2 out of 28 times (2/28 * 100 = 7%) or less.

Teachers who responded to question 33 indicated a common theme of frustration with SOL accountability. Survey results in paraphrased form denoted that high school teachers were concerned that they were being held responsible for the success of students who had historically failed SOLs for years. The responsibility of getting these students to pass their SOL test and acquire the needed verified credits to graduate with the lack of prior knowledge was daunting. High school teachers reported on the open ended
questionnaire item that they were given students randomly with no consideration from administration during the evaluation process on student’s ability level or prior testing results. Teacher reported that accountability based on SOL test scores was unfair for many reasons which included, ability grouping of students, grouping of students with behavior problems, and a student’s prior knowledge. Teachers indicated that the level of student that you teach greatly affected affect your SOL pass rates. Students in higher ability groups outperformed other students but many teachers struggled with raising test scores of students who were already outperforming their peers.

Multiple teachers commented that there was insufficient time spent on classroom observations. Teachers reported that observing two to three times a year for very brief time intervals does not give administration the proper information to evaluate a teacher’s performance. Veteran teachers believed that even less time is spent on their evaluation because of the concentration on new and struggling teachers. New teachers reported that while they were told they were doing things incorrectly they did not receive assistance to improve their performance. Teachers also reported not being observed for the number or length of time mandated by the teacher evaluation handbook. Teachers indicated that they did not blame administration for insufficient time spent in the classroom; they understood that the demands of the administrative jobs, such as dealing with student discipline, made the focus on evaluation nearly impossible. Additionally, teachers reported that central office staff provided no assistance to administration or the teaching staff in the process of improving classroom instruction.

A second theme that emerged was that the formal evaluation process provides teachers with very little meaningful information or feedback. Survey responses indicated
that the evaluation process was not meaningful because administration did not have the
time or resources to conduct meaningful observations and conferences that would
improve instruction. Teachers stated that evaluation feedback varied by evaluator with
some giving helpful tips and feedback and some providing virtually nothing in the form
of feedback. For some administrators, evaluation appeared to be a check off on the
administrative “to do” list and post observations conferences were simply a replay of
what happened in class during observation. Additionally, teachers reported that the
amount of observations conducted was insufficient to provide administration with a
proper view of a teacher’s performance. Teachers reported that administration was
simply going through the motions and that evaluation is a formality rather than an
experience of growth, learning, and improvement.

One response that was not repeated but worth noting was the statement, “I do my
job ‘meet the standard’ and do not try to go above or beyond which is reflected in my
performance evaluation.” This response encourages the study to question whether the
elevated scores on the performance evaluations where 65% of respondents report
receiving meets standard perpetuates mediocrity. A rating system where the majority of
teachers receive a meets standard rating could prompt teachers to do the minimum job
performance required instead of putting forth extra effort to receive exceeds standard. If
teachers are most likely to receive a meets standard rating unless extreme effort or failure
is shown, why would they go above and beyond?

It is important to note that while most comments from the teachers were critical of
the administrative process, including the frustration with SOL accountability, the
insufficient time spent on observation and conferences, and the lack of meaning in the
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

evaluation process, there were also a few comments that provided support for the current process commenting that the process was fine and should be left alone. It is also important to recognize that only 28 respondents answered question 33 out of 115 survey participants (24.3%). Therefore, the views of teachers answering question 33 represent a small portion of the total survey population.

Examination of the Research Questions

The purpose of this study was to investigate County X’s evaluation system to obtain an understanding of the impact that County X’s evaluation system has on teacher practice. Therefore, this study acquired teacher perspective and opinion through survey response and used this information to assess the evaluation systems influence on teacher practice.

This following research questions are examined in this section:

4. Are teachers who receive meaningful feedback more likely to use the results of performance evaluation for reflection?

H₄ - Teachers receiving meaningful feedback will be more likely to use the results of performance evaluation for reflection.

5. Are teachers who use the results of the performance evaluation for reflection more likely to report that the teacher evaluation system has a strong impact on their teaching practices?
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

H₅ - Teachers using results of the performance evaluation for reflection will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

6. Are teachers who receive meaningful feedback more likely to report that the teacher evaluation system has a strong impact on their teaching practices?

H₆ - Teachers receiving meaningful feedback will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

Table 5.18 shows the bivariate Pearson correlations comparing questions 21 and 27 that address hypothesis four, 27 and 31 to address hypothesis five, and 21 and 31 to address hypothesis six. The study used correlational analysis to address research question 4-6 listed above.

Table 5.18  Correlational Data to Address Research Questions 4-6.

<table>
<thead>
<tr>
<th>Question #21</th>
<th>Question #27</th>
<th>Question #32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback was useful for my professional development (N = 114)</td>
<td>Pearson Correlation Significance</td>
<td>1</td>
</tr>
<tr>
<td>Feedback from evaluator given during the post observation conference (N = 113)</td>
<td>Pearson Correlation Significance</td>
<td>.646*</td>
</tr>
<tr>
<td>Rate the overall impact of the evaluation on your teaching practices. (N=111)</td>
<td>Pearson Correlation Significance</td>
<td>.585*</td>
</tr>
</tbody>
</table>

NOTE: *Correlation is significant at the 0.01 level when a two tailed test is used
Research question number 4 asked: Are teachers who receive meaningful feedback more likely to use the results of performance evaluation for reflection? The correlation between question 21 and question 27 was positive, moderately strong, and significant $r = .646$, $p < .01$. The results supported the hypothesis which stated that teachers receiving meaningful feedback will be more likely to use the results of the performance evaluation for reflection.

Research has shown that teachers consistently seek more feedback from the evaluation process and believe that more feedback and follow up is needed (Rotheberg & Fenner, 1991; Kennedy, 2012). Teachers have also consistently stated that the purpose of evaluation should be for reflection and professional growth (Peterson & Comeaux, 1990; Feeney, 2007). The findings for research question number four are consistent with previous studies which have demonstrated that teachers seek meaningful feedback, value feedback as a necessary component in evaluation, and believe that evaluation should be for reflection and professional growth (Rotheberg & Fenner, 1991; Kennedy, 2012; Peterson & Comeaux, 1990; Feeney, 2007).

Research question number 5 asked: Are teachers who use the results of the performance evaluation for reflection more likely to report that the teacher evaluation system has a strong impact on their teaching practices? The correlation between questions 27 and question 32 was found to be positive, moderately strong, and significant $r = .593$, $p < .01$. The results supported the hypothesis which stated that teachers using the results of the performance evaluation for reflection will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.
Peterson & Comeaux (1990) found that teachers believed evaluation should facilitate a teacher’s reflection on his or her classroom practice, which should, in turn improve a teacher’s classroom skills and promote their professional growth. Additionally, research has shown that teachers believed that if they reflect on their performance evaluation, the result will positively impact their effectiveness as a teacher (Peterson & Comeaux, 1990). Unfortunately, research has also shown that when quality feedback and follow up are missing little reflection occurs which causes teachers to perceive evaluations not to be meaningful or worthy of reflection (Blumberg, 1976; Mahar & Strobert, 2010; Derrington 2011). Therefore, it was hypothesized that if teachers actually reflect upon their performance evaluations they will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices. This conclusion was supported by the findings of this research investigation.

Research question number 6 asked: Are teachers who receive meaningful feedback more likely to report that the teacher evaluation system has a strong impact on their teaching practices? The correlation between questions 21 and question 32 was found to be positive, moderately strong, and significant $r = .585, p < .01$. The results supported the hypothesis which stated that teachers who received meaningful feedback will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

The research has shown that when quality feedback and follow up is missing from evaluation procedures, teachers perceive evaluation to be an ineffective formality (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011). Research has also indicated that an evaluation instrument would only be effective if teachers actually view
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

The instrument as fair and valid (Peterson & Comeaux, 1990; Kyriakides, Demetriou, & Charalambous, 2006; Tyler, 2005). In order for the instrument to be valid, and therefore effective, studies have found that the instrument must provide quality feedback and administrators must follow up with the teacher after the rating (Rothberg & Fenner, 1991; Kennedy, 2012). The results of this investigation support the research by finding a strong positive correlation to exist between teachers who report receiving meaningful feedback and those that report that the evaluation system has a strong impact on their teaching performance.

**Difference in Means, Correlational and Multivariate Analysis**

This research investigation has established that teachers who receive meaningful feedback reflected on their teaching practice and reported a larger overall impact from the evaluation process on their teaching performance. This section presents results about factor associated with teachers’ perceptions of the overall quality of the evaluation and the impact of the evaluation on their teaching practice using difference of means tests (t-test), correlations, and multiple regressions.

This study demonstrated that the mean differences between the overall quality of the evaluation (mean 3.45) and the overall impact of the evaluation on teaching practice (mean 3.10) were statistically significant (T= 4.73, n= 112, p= .000). Respondents rated the quality of the evaluation higher than the impact that the evaluation had on actual teaching practice and the difference reflected is significant. It is possible that the quality of the evaluation was higher due to 98.2% of respondents reporting they received a meets or exceeds standards summative rating. Teachers were satisfied with their
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

summative evaluation scores but did not believe that the evaluation process impacted their classroom instruction.

This investigation conducted bivariate correlations on the items contained in the four survey sections validity, evaluation feedback, multiple sources of data used, and reflection on practice to examine items contributing to teacher’s perception of the overall quality of the evaluation and the overall impact of the evaluation on their teaching practice.

The results from the bivariate correlations calculated for questions 8-14 of the validity section are listed in Table 5.19. The validity section revealed that Q11 (the number of observations and times spent conducting classroom observations $r = .659, p < .01$) had the highest correlation with the rating of overall quality of the evaluation. The next highest correlation with the rating of overall quality of the evaluation was Q14 (the teacher evaluation instrument used to rate my performance is fair and valid $r = .643, p < .01$). The strongest correlation with the rating of overall impact of the evaluation on teaching practice was with Q13 (the information collected in the documentation log provides evidence of several performance standards; these items are a valid way of providing a comprehensive portrait of my work $r = .505, p <.01$) followed by Q11 (the number of observations and time spent conducting classroom observations $r = .471, p < .01$).

The number of observations and time spent conducting classroom observations was significant to both the overall quality and the overall impact of the evaluation process. Evaluation fairness and validity correlated much more strongly with the quality of the evaluation process ($r = .643, p < .01$) than it did with the impact the
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

evaluation process had on classroom teaching \( r = .452, p < .01 \). Once again the higher correlation may be influenced by the elevated summative performance scores (98.2% teachers report receiving meets or exceeds standard).
Table 5.19 Bivariate Correlations Survey Section 2: Validity

<table>
<thead>
<tr>
<th></th>
<th>Rate the Overall Quality of the Evaluation (Q31)</th>
<th>Rate the Overall Impact of the evaluation on your teaching practice (Q32)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Significance</td>
</tr>
<tr>
<td>The results of my summative evaluation accurately portrayed my performance. (Q8)</td>
<td>111</td>
<td>.000</td>
</tr>
<tr>
<td>Clear examples were given to justify the ratings I received on my summative evaluation. (Q9)</td>
<td>111</td>
<td>.000</td>
</tr>
<tr>
<td>The rating I received on my summative evaluation did not reflect bias from my evaluator. (Q10)</td>
<td>110</td>
<td>.000</td>
</tr>
<tr>
<td>The number of observations and times spent conducting classroom observations including pre and post observation conferences allows for an accurate prediction of my teaching performance. (Q11)</td>
<td>112</td>
<td>.000</td>
</tr>
<tr>
<td>The measurement of student performance that is evaluated during the goal setting process provided an accurate picture of my ability to impact student learning. (Q12)</td>
<td>112</td>
<td>.000</td>
</tr>
<tr>
<td>The information collected in the documentation log provides evidence of several performance standards. These items are a valid way of providing a comprehensive portrait of my work. (Q13)</td>
<td>111</td>
<td>.000</td>
</tr>
<tr>
<td>The teacher evaluation instrument used to rate my performance is fair and valid. (Q14)</td>
<td>112</td>
<td>.000</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the .01 level (2 tailed)
Further exploration through bivariate correlation into evaluation feedback (questions 15-22) is shown in Table 5.20. Both of the dependent variables Q31 (overall quality) and Q32 (impact of the evaluation on teacher practice) had the highest correlation with question 22 (feedback was meaningful and assisted me to improve my classroom instruction). The lowest correlational value for both of the dependent variables Q31 and Q32 was timing of the feedback during evaluation. Therefore, meaningful feedback that assisted teachers in improving their classroom instruction had the greatest impact on the rating of overall quality of the evaluation and the overall impact of the evaluation on teacher performance.
Table 5.20 Bivariate Correlations Survey Section 3: Evaluation Feedback.

<table>
<thead>
<tr>
<th>Rate the Overall Quality of the Evaluation (Q31)</th>
<th>Rate the Overall Impact of the Evaluation on Your Teaching Practice (Q32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Pearson Correlation</td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
</tr>
<tr>
<td>Amount of Information received (Q15)</td>
<td>112</td>
</tr>
<tr>
<td>Depth of information received (Q16)</td>
<td>111</td>
</tr>
<tr>
<td>Quality of the ideas and suggestions contained in the feedback (Q17)</td>
<td>112</td>
</tr>
<tr>
<td>Specificity of information provided (Q18)</td>
<td>112</td>
</tr>
<tr>
<td>Timing of the feedback (Q19)</td>
<td>111</td>
</tr>
<tr>
<td>Application of information toward student achievement (Q20)</td>
<td>112</td>
</tr>
<tr>
<td>Feedback was useful for my professional development (Q21)</td>
<td>112</td>
</tr>
<tr>
<td>Feedback was meaningful and assisted me to improve my classroom instruction (Q22)</td>
<td>112</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the .01 level  
*Correlation is significant at the .05 level

Table 5.21 displays the bivariate correlation of questions 23-26 multiple sources of data. The dependent variable Q31 (rate the overall quality of the evaluation) showed the strongest correlation ($r = .604$, $p < .01$) with Q25 (student survey summary) and the dependent variable Q32 (rate the overall impact of your evaluation on your teaching practice) had the strongest correlation ($r = .602$, $p < .01$) with Q24 (examination of
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district artifacts in the documentation log). Therefore, the greatest impact on quality came from teacher reflection on student survey data. County X’s student survey stated that the purpose of the survey was to allow students to provide teachers with input into how the class might be improved (Carnot et. al., 2007). However, the survey actually gives teachers a numerical rating (see Appendix A student survey and student survey summary). Teachers reflected upon survey results and summarized their thoughts in the student survey summary. A teacher’s perception of the value of this instrument would be influenced by the scores the teacher received on the student surveys.

The greatest influence on the overall impact of a teacher’s evaluation on their teaching practice was the examination of artifacts in the documentation log. This documentation log included listing of parent contacts, professional development, student test scores and academic progression, syllabus, lesson plans etc. The documentation log allowed the teacher to present a complete picture of what happens in the classroom to administration during the post observation conference.

Data reviewed revealed a distinction between quality Q31 (do I agree with the rating I have received) and impact Q32 (does this practice improve my teaching).
Table 5.21 Bivariate Correlation of Survey Section 4: Multiple Sources of Data

<table>
<thead>
<tr>
<th></th>
<th>Rate the Overall Quality of the Evaluation (Q31)</th>
<th>Rate the Overall Impact of the evaluation on your teaching practice (Q32)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Pearson Correlation</td>
</tr>
<tr>
<td>Observation of your classroom performance (Q23)</td>
<td>112</td>
<td>.268**</td>
</tr>
<tr>
<td>Examination of artifacts in your documentation log (Q24)</td>
<td>110</td>
<td>.560**</td>
</tr>
<tr>
<td>Student Survey Summary (Q25)</td>
<td>111</td>
<td>.604**</td>
</tr>
<tr>
<td>Goal setting to assess student achievement (Q26)</td>
<td>111</td>
<td>.480**</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the .01 level (2 tailed)

The data for the last survey section (reflection on practice) analyzed by bivariate correlation is shown in Table 5.22. Both of the dependent variables Q31 (overall quality) and Q32 (impact of the evaluation on teacher practice) had the highest correlation with question 27 (feedback from evaluator given during post observation conference).

Feedback is an essential component that affects both the rating of quality and impact of the evaluation process.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 5.22 Bivariate Correlation of Survey Section 5: Reflection on Practice

<table>
<thead>
<tr>
<th>Rate the Overall Quality of the Evaluation (Q31)</th>
<th>Rate the Overall Impact of the Evaluation on Your Teaching Practice (Q32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Pearson Correlation</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Feedback from evaluator given during post observation conference (Q27)</td>
<td>112</td>
</tr>
<tr>
<td>Collection of artifacts in your documentation log (Q28)</td>
<td>112</td>
</tr>
<tr>
<td>Completion of the student survey and student survey summary (Q29)</td>
<td>112</td>
</tr>
<tr>
<td>Data analysis to assess student achievement as part of the goal setting process (Q30)</td>
<td>112</td>
</tr>
</tbody>
</table>

Note: ** Correlation is significant at the .01 level (2 tailed)

In summary, the bivariate correlation revealed that Q31 (overall quality) and Q32 (impact of the evaluation on teacher practice) were both strongly influenced by Q27 (feedback from evaluator during post observation conference), Q22 (meaningful feedback to assist me to improve instruction) and Q11 (the number of observations and times spent conducting classroom observations including pre and post observation conferences). The overall quality of the evaluation and overall impact of the evaluation on teacher practice were both clearly influenced by feedback and time spent performing observations and observation conferences. However, the two dependent variable clearly differed when Q31
(overall quality) was associated with Q14 (instrument being fair and valid) and Q25 (student survey summary) while Q32 (overall impact on teachers practice) was correlated with Q13 (information collected in documentation log) and Q24 (examination of artifacts in the documentation log). Data reviewed revealed a distinction between quality Q31 (do I agree with the rating I have received) and impact Q32 (does this practice improve my teaching).

To further investigate the difference between the scores on the quality of the evaluation and the impact that the evaluation had on the actual teaching practice, multiple linear regressions were conducted. Specifically, this study investigated the impact of the perceived validity of the performance evaluation process index ($X_1$), the quality of the evaluation feedback index ($X_2$), the use of multiple sources of data index ($X_3$), and the extent to which teachers reflect on practice index ($X_4$) on the overall rating for quality of evaluation and the overall rating of impact of the evaluation on teaching practice.

$$Y=b_0 + b_1X_1 +b_2X_2 +b_3X_3+ b_4X_4$$

Appendix F describes the survey questions that were mapped to indices and descriptive statistics for $X_1$, $X_2$, $X_3$, and $X_4$. The reliability coefficients support combining the criteria to represent the overall construct of performance validity $X_1$ (alpha=.883), evaluation feedback $X_2$ (alpha=.935), multiple sources of data $X_3$ (alpha = .725), and reflection on practice $X_4$ (alpha = .770). Johnson and Christensen (2012) verify that to show reliability the coefficient alpha for research purposes should be greater than .70.
The findings for the OLS regression analysis are presented in Table 5.23. Overall the model predicted 69% \((R^2 = .69)\) of the variance in the dependent variable (rate the overall quality of the evaluation Q31). The overall model was significant \((F= 55.53 \, p= .000)\). Two of the four indices, validity \((t= 5.2, \, p=.000)\) and evaluation feedback \((t=3.4, \, p=.001)\), were significant predictors of the rating of the overall quality of the evaluation.

As shown in Table 5.23, the model predicted 56% \((R^2 = .56)\) of the variance in the dependent variable (rate the overall impact of the evaluation on your teaching practice Q32). This model was also significant \((F= 32.49, \, p= .000)\) at the .01 level. Two of the four indices evaluation feedback \((t=3.51, \, p=.001)\) and multiple sources of data \((t=2.71, \, p=.008)\), were significant predictors of overall impact of the evaluation on teaching practice.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Table 5.23 OLS Regression Results of Survey Data.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate the Overall Quality of the Evaluation (Q31)</td>
<td>B</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Validity</td>
<td>.464</td>
<td>.089</td>
</tr>
<tr>
<td>Evaluation Feedback</td>
<td>.123</td>
<td>.114</td>
</tr>
<tr>
<td>Multiple Sources of Data</td>
<td>.340</td>
<td>.099</td>
</tr>
<tr>
<td>Reflection on Practice</td>
<td>.146</td>
<td>.089</td>
</tr>
</tbody>
</table>

N= 106
R² = .687, Adjusted R²= .675, F value = 55.53**
**Significant at .01

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Unstandardized Coefficient</th>
<th>Standardized Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate the Overall Impact of the Evaluation on Your Teaching Practice (Q32)</td>
<td>B</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Validity</td>
<td>.098</td>
<td>.116</td>
</tr>
<tr>
<td>Evaluation Feedback</td>
<td>.456</td>
<td>.130</td>
</tr>
<tr>
<td>Multiple Sources of Data</td>
<td>.404</td>
<td>.149</td>
</tr>
<tr>
<td>Reflection on Practice</td>
<td>.102</td>
<td>.116</td>
</tr>
</tbody>
</table>

N=106
R²=.563, Adjusted R²=.545, F value = 32.49**
**Significant at .01

Consistent with the bivariate analysis presented earlier, the multiple regression results showed that evaluation feedback had a significant impact on both of the
dependent variables (rate the overall quality of the evaluation Q31 and rate the overall impact of the evaluation on your teaching practice Q32). Research has indicated that teachers want more feedback and follow up during the evaluation process to grow professionally (Rothberg & Fenner, 1991; Kennedy, 2012; Mahar & Strobert, 2011; Peterson & Comeaux, 1990). Additionally, studies have shown that teachers seek meaningful feedback, value feedback as a necessary component in evaluation, and believe that evaluation should be for reflection and professional growth (Rothberg & Fenner 1991; Kennedy 2012; Peterson & Comeaux, 1990; Feeney, 2007). Results of this study supported previous research and the study’s hypothesis: teachers who receive meaningful feedback will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

Researchers have also shown that quality feedback is often a missing component from the teacher evaluation process (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2011). The data from the evaluation feedback section (survey questions 15-20) supported prior research. Data from this study revealed that teachers were receiving a large amount of feedback that was specific and timely; however, the information needs more depth, quality, and relevance toward student achievement for the post observation conference to actually influence a teacher’s professional development or assist teachers with improvement of classroom instruction. Additionally, this study showed that the average rating for the overall impact of the evaluation process on a teachers practice was a 3.10 indicating that the highest level of respondents 32% said that the evaluation had an average impact on their teaching practice. The rating Q32 (overall impact of the
evaluation on your teaching practice) was impacted by the lack of quality feedback given on performance evaluations in County X.

Additionally, the perception of impact of performance evaluations on teaching practice Q32 was influenced by the collection of multiple sources of data, especially examination of artifacts in the documentation log. Prior research revealed that teachers find evaluations that used multiple sources of data collection such as documentation logs, peer observations, stakeholder surveys, and student achievement data, along with multiple classroom observations were more helpful in providing the desired feedback to improve teacher instruction and professional growth (Peterson & Comeaux, 1990; Mahar & Strobert, 2010; Rothberg & Fenner, 1991).

Therefore, to increase a teacher’s perception of the overall impact that the evaluation process has on his or her practice, administrators must give quality feedback that is meaningful and relevant during post observation conferences. They must also increase time spent analyzing, reviewing, and discussing other items included in multiple sources of data such as student surveys, collection of evidence in the documentation log, and goal setting to assess student progress. Post observation discussion must include a review of multiple sources of data so that the information is meaningful to improve a teacher’s classroom instruction.

Teachers’ perception of the overall quality of the evaluation Q31 was also influenced by the index of validity. Prior research indicated, based on teacher opinion, that evaluation instruments were effective only if they were perceived as fair and valid (Peterson & Comeaux 1990; Kyriakides, Demetriou, & Charalambous, 2006; Tyler 2005). When teachers did not view the performance evaluation instrument as a valid
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

tool for professional growth their perception became that they only needed to pass the teacher performance evaluation, not actually learn and improve from the experience (Peterson & Comeaux, 1990). Teachers in County X (98.2% reported receiving meets or exceeds the standard) were performing at or above the standard on their performance evaluations. The combination of the index of validity effecting Q31 (quality of the evaluation) rather than Q32 (impact of the evaluation on teacher practice) and the high level of teachers receiving meet or exceeds standard could explain the significant difference observed between the overall quality of the evaluation and the overall impact of the evaluation on teacher practices.

**Summary**

A summary of the findings reveal that the majority of teachers received meets or exceeds standard performance score. In the summative performance evaluation 72.4% of teachers received meets or exceeds the standard and on the holistic performance score 65.5% of the teachers received meets or exceeds the standard. This investigation did not find a large discrepancy between the summative performance evaluation score and the holistic performance score as expected. Conversely, this investigation actually found a positive, significant correlation between the summative evaluation performance score and the administrative holistic performance score, indicating that the lack of differentiation in performance scores was not attributed to the inability of researchers to give strong negative feedback. The other possible explanation to these findings that must be considered is that neither the summative performance evaluation score nor the holistic performance score required administration to give strong negative feedback. This explanation is a possibility since the summative performance evaluation does not require the administration to give an overall summative score. Scores were given for
each of the seven performance categories but an overall summative score was not present on the summative evaluation document. Therefore, administration could avoid discussing a teachers overall performance rating.

Findings indicated that the largest percentage of teachers receiving need improvement or unsatisfactory was in the category of student achievement with 38.9% of teachers receiving a rating that was less than meeting the standard. However, if teachers only received one needs improvement rating the overall summative evaluation score would still be meets standard or higher. A teacher must receive one unsatisfactory rating or two needs improvement ratings for the teacher to be placed on a plan of improvement and have an overall rating below meets standard. Therefore while teachers were being marked lower in the student achievement category, this had little to no effect on the overall summative performance score.

Additionally, the summative evaluation score showed a significant positive correlation with the student performance score and the holistic performance assessment score did not have a significant correlation with the student performance score. These results indicated that a standards based assessment instrument that uses multiple sources of data is better at predicting teacher performance based on student achievement than administrative opinion alone. This investigation confirmed that the summative performance evaluation was a better tool of assessing teacher’s performance based on student achievement than the administrative holistic performance score.

Similarly, the majority of teachers (98.2%) of the 115 teachers surveyed reported that they received a meets or exceeds standard on the summative performance evaluation. The majority of teachers agreed that the summative evaluation was an
accurate portrayal of their performance, but the tools used such as observation, goal setting process, and the collection of evidence in the documentation log were not valid ways of assessing their performance. It is not completely understood whether agreement with the summative evaluation procedure comes from the score itself or the actual process since the scores reported are almost exclusively meets or exceeds the standard.

Results did indicate that time spent on observations and clear examples provided to justify performance ratings had a significant impact on a teacher’s perception of the quality of the evaluation. Additionally, the use of multiple sources of data (observation, documentation log, student survey, goal setting process) was shown to affect the overall impact of the evaluation on teaching practices. Therefore the lack of sufficient time spent conducting these activities may be the reason they are not seen as valid assessments by teachers.

Another possible reason teachers do not view these documents as valid is the quality of feedback and discussions that teachers are receiving during post observation conferences and during the review of documents that are considered in multiple source of data. Findings indicated that feedback on post observation conferences was specific and timely but greater depth, quality, and relevance was needed for this to be a meaningful process that promoted professional growth and classroom improvement. Additionally, results indicated that evaluation feedback, specifically meaningful feedback that assisted teachers to improve classroom instruction, was found to have a significant impact on the overall quality of the evaluation as well as the overall impact of the evaluation on a teachers practice. Thus to improve the overall quality and impact
of the evaluation process in County X administration must improve the quality of feedback provided to teachers during the evaluation process.

The need for greater, depth, quality, and relevance of feedback during the post observation conference was also a finding in the open ended response section where 28 of the 115 teachers replied. Teachers reported frustration with SOL accountability, believed that there is insufficient time spend in classroom observations, and reported that they regard the evaluation process as a formality rather than a meaningful process. More time must be spent on observation and meaningful, relevant, discussions must happen during the post observation conference in order for teachers to view this process as a meaningful learning experience that promotes professional growth.

Finally, investigation of research questions four through six indicated that when teachers were given meaningful feedback they used this information to reflect on their professional growth, when reflection occurs teachers reported that the performance evaluation had a strong impact on their teaching practice. One problem observed in County X was that administrators were spending insufficient time on the evaluations and not all teachers were receiving meaningful feedback required to facilitate reflection and subsequently improvement.
Discussion and Conclusions

Chapter 6

Introduction

In chapter 6, a discussion the results of and conclusions determined from this study is presented. The chapter begins with a review of the significance of the study which includes the problem and purpose of the research investigation. Next, the chapter presents an overview of the significant findings which includes methodology followed and a summary of findings by research question. The chapter concludes with a presentation of research findings, conclusions and recommendations for County X, limitations of the study, and recommendations for further research.

Significance of the Study

Substantial funding is being provided from both the state and federal government for school reform that links teacher pay for performance with student achievement. The pool of money available to schools to implement reforms comes with significant concerns over how school systems will attempt to link teacher evaluations with student achievement gains. Obtaining a valid estimate of a teacher’s actual contribution to the student learning is a daunting task for school divisions (Steele, Hamilton, & Stecher, 2010). It is therefore critically important for school divisions to thoroughly investigate their performance evaluation system to insure that it is a fair and effective method to evaluate teacher performance and measure student achievement.
This study investigated County X’s evaluation system to find if the teacher performance evaluation system influenced student achievement. Additionally, this performance evaluation research investigated the impact that County X’s evaluation system had on a teacher’s instructional practices. This performance evaluation research contributed to existing research because it built upon previous studies and expanded their approach. This performance evaluation research included all aspects of teacher summative performance evaluation, by including an administrative holistic performance rating, and by investigating the influence on teacher practice using teacher perspective and opinion. This research study assessed the efficiency of the current teacher evaluation practice and used the data and conclusions from this research to recommend improvements.

Overview of Significant Findings

This section begins with a description of the methodology used in this study and then provides a summary of findings for the three research questions addressed in Part 1: Performance Evaluation Systems Correlation between Teacher Performance and Student Achievement – In depth Study of a Subset of Teachers in County X and the three research questions addressed in Part 2: The Impact of Performance Evaluations on Teacher Instructional Practice-Survey of High School Teachers in County X.

Methodology

This program evaluation of County X’s teacher evaluation system collected data from Algebra I, Algebra II, and Geometry teachers’ summative evaluations from the 2011-2012 and the 2012-2013 school years making this research a cross-sectional
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district.

Descriptive study. The sample reviewed summative evaluation documents of 25 SOL math teachers. The determination of whether County X’s teacher evaluation system could accurately predict student achievement was obtained by determining if a strong positive correlation existed between teachers’ summative performance evaluation scores and student performance data. This study also examined the link between administrative holistic teacher performance assessment and the student performance data, as well as the correlation between the administrative holistic performance assessment and teachers’ summative evaluation performance scores.

Additionally, this research investigation examined the impact that County X’s evaluation system had on teacher practice. To assess teacher opinion 234 surveys were distributed to teachers in the three high schools contained in County X. This survey obtained a broad perspective on the impact that the evaluation system has on teacher practice. This investigation used t-test, bivariate correlation, and multivariate analysis to investigate items contributing to teacher’s perception of the overall quality of the evaluation and the overall impact of the evaluation on their teaching practice.

**Summary of Findings by Research Questions**

Part I of this research study examined three research questions using data collected from a subset of mathematics teachers in County X. These three questions investigated whether County X’s teacher evaluation system could accurately predict student achievement in mathematics.

Part 1: Performance Evaluation Systems Correlation between Teacher Performance and Student Achievement – In-depth Study of a Subset of Teachers in County X.
1. Do the scores on the teacher summative performance evaluation correlate to the student performance score?

H₁ – Teacher summative performance evaluation score will not be correlated with student performance score. Not Supported

The correlation between summative performance evaluation and student performance score was found to be positive and significant r (29) = .395, p=.034. The results refuted the hypothesis that teacher summative evaluation scores would not be correlated with the student performance score.

2. Do administrative holistic performance assessments correlate to the student performance score?

H₂ – Administrators holistic performance assessment score will be correlated positively with the student performance scores. Not Supported

The correlation between administrative holistic performance assessments and the student performance score was positive r (29) = .265 but not significant p = .165. These results refuted the hypothesis that the administrative holistic performance assessment score would be correlated positively with the student performance score.

3. Do the teacher summative evaluation performance scores correlate to the administrative holistic performance assessment score?

H₃ – Teacher summative performance score will not be correlated with administrative holistic performance assessment score. Not Supported

From the collection of data on 25 mathematics teachers, results showed that the teacher summative evaluation performance scores and the administrative holistic performance assessment of teacher performance were significantly related, r (25) = .76, p
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

< .01. The results did not support hypothesis 3; teacher summative performance scores were not correlated with administrative holistic performance scores.

This investigation did not find a significant positive correlation between administrative holistic performance score and the student performance score. However, it did find a positive and significant correlation between teacher summative performance evaluation score and the student performance score $r = .395$, $p = .034$. Prior research reviewed found studies that investigate the link between teacher performance evaluations and student achievement only yielded weak correlational values to represent the relationship (Kane et al., 2010; Kimball et al., 2004; Milanowski, 2004; Washington, 2011). The significant correlation between teacher summative performance evaluation and the student performance score validates County X’s teacher evaluation system.

This research is consistent with the studies that examined the relationship using a standards based approach that included multiple sources of data, produce a more significant link between student achievement and teacher performance (Hinchey, 2010; Jacob & Lefgren, 2008; Kane et al., 2010; Kimball et al., 2004; Stronge et al., 2011). Therefore, this research provided additional support for the use of a standard based approach that uses multiple sources of data as a stronger link to student performance than a holistic administrative appraisal of a teacher’s performance. Although this study showed that the summative performance scores to be a better predictor of student achievement, the significant positive correlation between the summative performance scores and the holistic performance assessment scores also showed that the holistic performance evaluation scores holds value in the evaluation process.
Part 2 of this research study examined three research questions using data collected from the distribution of surveys in three high schools in County X. These three questions investigated the impact that the evaluation system had on teacher practice.

Part 2: The Impact of Performance Evaluations on Teacher Instructional Practice – Survey of High School Teachers in County X

4. Are teachers who receive meaningful feedback more likely to use the results of performance evaluation for reflection?

H₄ - Teachers receiving meaningful feedback will be more likely to use the results of performance evaluation for reflection. **Supported**

The correlation between question 21 (Feedback was useful for my professional development) and question 27 (Report the extent the following sources of performance information caused you to reflect and improve your teaching practice - Feedback from evaluator given post observation conference) was moderately strong, positive, and significant $r = .646, p < .01$. The results supported the hypothesis which stated that teachers receiving meaningful feedback will be more likely to use the results of the performance evaluation for reflection.

5. Are teachers who use the results of the performance evaluation for reflection more likely to report that the teacher evaluation system has a strong impact on their teaching practices?

H₅ - Teachers using results of the performance evaluation for reflection will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices. **Supported**
The correlation between question 27 (Report the extent the following sources of performance information caused you to reflect and improve your teaching practice - Feedback from evaluator given post observation conference) and 32 (Rate the overall impact of the evaluation on your teaching practices) was moderately strong, positive, and significant $r = .593, p < .01$. The results supported the hypothesis; teachers using the results of the performance evaluation for reflection will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

6. Are teachers who receive meaningful feedback more likely to report that the teacher evaluation system has a strong impact on their teaching practices?

$H_6$: Teachers receiving meaningful feedback will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

Supported

The correlation between question 21 (Feedback was useful for my professional development) and question 32 (Rate the overall impact of the evaluation on your teaching practices) was moderately strong, positive, and significant $r = .585, p < .01$. The results supported the hypothesis which stated that teachers who received meaningful feedback will be more likely to report that the teacher evaluation system has a strong impact on their teaching practices.

Research indicated that teachers seek meaningful feedback, value feedback as a necessary component in evaluation, and believe that evaluation should be for reflection and professional growth (Rothberg & Fenner, 1991; Kennedy, 2012, Peterson & Comeaux, 1990; Feeney, 2007). Additionally, research showed that teachers believe that
evaluation should facilitate a teacher’s reflection on his or her classroom practice, which should in turn improve a teacher’s classroom skills (Peterson & Comeaux, 1990). The results of this investigation supported the research by finding a strong positive correlation to exist between teachers who report receiving meaningful feedback and those that report that the evaluation system has a strong impact on their teaching performance.

Unfortunately, when quality feedback and follow up are missing little reflection occurs which causes teachers to perceive evaluations not to be meaningful or worthy of reflection (Blumberg, 1976; Mahar & Strobert, 2010; Derrington, 2010). Research showed that when quality feedback and follow up are missing from evaluation procedures, teachers perceived evaluation to be an ineffective formality (Blumberg, 1976, Mahar & Strobert, 2010; Derrington, 2010). Feedback is the key factor in the perceptions that teachers have in regard to the impact of the evaluation system on their teaching practice.

**Research Findings**

Overall findings from the summative performance evaluation scores, administrative holistic performance scores, and student achievement scores in County X revealed that the summative evaluation score \( r(29) = .395, p = .034 \) was a stronger predictor of student achievement than the holistic performance score \( r(29) = .265, p = .165 \). This research investigation provided support that the summative performance evaluation procedure in County X which used a standards based approach and included multiple sources of data was a stronger predictor of student achievement than the administrative holistic appraisal of teaching performance.
Although research revealed that a strong positive significant correlation existed between the holistic performance assessment score and the summative evaluation score \((r (25) = .76, p< .01)\) differences between the two categories were revealed. The data from the categories of student achievement \((r= .347, p= .089)\) and assessment \((r= .200, p=.337)\) did not provide a significant correlation with the holistic performance assessment score. Additionally, descriptive statistics revealed that the largest percentage of teachers (38.9 %) receiving needs improvement or unsatisfactory was in category of student achievement. Therefore, data showed that the summative evaluation assessment scores deviated from the holistic performance assessment in the areas of student achievement and assessment.

Furthermore, by examining a paired t-test, this study demonstrated that the mean differences between the administrative holistic performance assessment score and the summative evaluation score were statistically significant \((T= -2.366, n=25, p=.03)\). The study revealed that mean scores were significantly higher on the summative performance assessment than the holistic performance assessment in the categories of professionalism and instructional delivery. This data were supported by descriptive statistics which indicated that the highest percentage of teachers receiving exceeds standard was in the category of instructional delivery and professionalism.

Although the summative evaluation instrument in County X produced a significant correlation with the student performance score, teachers received the highest ratings in the category of instructional delivery and the lowest ratings in student achievement on the summative performance assessment. It is unexpected that the categories of instructional delivery and student achievement would be in opposition to
each other. The data also revealed additional concerns with 76% of teachers receiving ratings of meets or exceeds standard. These elevated performance scores and lack of variance in performance data (variance of summative performance score = .14) did not coincide with the mean of the student achievement score being 398, which is below a passing score of 400.

Similarly, the majority of teachers (98.2%) of the 115 teachers surveyed reported that they received a meets or exceeds standard on the summative performance evaluation. The majority of teachers agreed that the summative evaluation was an accurate portrayal of their performance, but the tools used such as observation, goal setting process, and the collection of evidence in the documentation log were not valid ways of assessing their performance. It is not completely understood whether agreement with the summative evaluation procedure comes from the score itself or the actual process since the scores reported are almost exclusively meets or exceeds the standard.

Results indicated that time spent on observations and clear examples provided to justify performance ratings impacted a teacher’s perception of the quality of the evaluation. Additionally, teachers specified that the use of multiple sources of data (observation, documentation log, student survey, goal setting process) affected the overall impact of the evaluation on teaching practices. Therefore the lack of sufficient time spent conducting these activities may be the reason they are not seen as valid assessments by teachers.

Another possible reason teachers do not view these documents as valid is the quality of feedback and discussions that teachers received during post observation
conferences and during the review of documents that are considered in multiple source of data. Findings indicated that feedback on post observation conferences were specific and timely but greater depth, quality, and relevance was needed for this to be a meaningful process that promoted professional growth and classroom improvement. Additionally, results indicated that evaluation feedback, specifically meaningful feedback that assisted teachers to improve classroom instruction, provided a significant impact on the overall quality of the evaluation as well as the overall impact of the evaluation on a teachers practice. Thus to improve the overall quality and impact of the evaluation process in County X administration must improve the quality of feedback provided to teachers during the evaluation process.

In the open ended response section of the survey where 28 out of the 115 teachers responded, teachers once again specified a need for greater, depth, quality, and relevance of feedback during the post observation conference. Teachers also reported frustration with SOL accountability, felt that there was insufficient time spent in classroom observations, and reported that they regard the evaluation process as a formality rather than a meaningful process. More time must be spent on observation and meaningful, relevant, discussions must happen during the post observation conference in order for teachers to view this process as a meaningful learning experience that promotes professional growth.

Finally, this investigation found that when teachers were given meaningful feedback they used this information to reflect on their professional growth, when reflection occurred teachers reported that the performance evaluation has a strong impact on their teaching practice. The problem that occurred in County X was that
sufficient time was not being spent during the evaluation process and not all teachers were receiving the meaningful feedback required to facilitate reflection so that improvement occurred.

**Conclusions and Recommendations for County X**

County X’s summative performance evaluations have a moderately strong positive correlation with the student performance score. The significant correlation between teacher summative performance evaluation and the student performance score validates County X’s teacher evaluation system. County X’s standards based approach that included the review of multiple sources of data has produced a link between student achievement and teacher performance.

**Recommendation #1**

County X should continue the practice of using the summative performance instrument which contains seven performance standards and a review of multiple sources of data to evaluate teacher performance.

This research investigation found a strong positive correlation between the summative performance evaluation score and the holistic performance assessment score. Further investigation of the data identified deviations between the two scores in student achievement and assessment with the summative performance evaluation being the better predictor of student performance. Additionally, the summative performance evaluation showed higher scores in the standards of professionalism and instructional delivery.

This investigation noted as a concern the lack of relationship shown between the standard
of achievement and instructional delivery in the summative performance evaluation score.

**Recommendation #2**

County X should review the standards and corresponding observable behaviors (performance indicators) for the standard of instructional delivery. County X must ensure that indicators are clearly defined and properly linked to classroom behaviors that are predictive of student achievement.

In County X, performance scores were elevated with 76% of teachers receiving a rating of meets or exceeds the standard. These elevated performance scores and lack of variance in performance data (variance of summative performance score = .14) did not coincide with the mean of the student achievement score being 398 which is below passing score of 400.

**Recommendation #3**

Even though this research investigation found a moderately strong correlational link between teacher performance and student achievement, it is important for County X to continue to strengthen the connection between a teacher’s summative performance score and student performance data. In order for County X to surpass the current link established they must investigate the elevation and lack of variation in performance scores. This process would start by reviewing performance indicators for each of the seven standard to ensure that a clear distinction is made between the observed behaviors leading to a rating of exceeds standard, meets standard, needs improvement, and unsatisfactory.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Teachers in County X reported that tools used such as observation, goal setting process, and the collection of evidence in the documentation log were not valid ways of assessing their performance. Findings indicated that feedback on post observation conferences were specific and timely but greater depth, quality and relevance was needed for this to be a meaningful process that promoted professional growth and classroom improvement. Teachers identified meaningful feedback as a critical component that provided relevance to observation and the collection of evidence in the documentation log. Post observation conferences must contain meaningful, relevant discussions in regard to observed classroom behaviors as well as the progression of student achievement and the collection of evidence in the documentation log.

**Recommendation #4**

To improve the overall quality and impact of the evaluation process in County X; administration must improve the quality of feedback provided to teachers during the evaluation process. Sufficient time must be spent reviewing documents from documentation log during post observation conference. More time must be spent on observation and meaningful, relevant, discussions must occur during the post observation conference in order for teachers to view this process as a meaningful learning experience that promotes professional growth.

**Recommendation #5**

It is a recommendation of this study that County X conduct training for administrators and evaluators to facilitate understanding of the observed characteristics that should be exemplified for each standard and performance rating. It is crucial for all evaluators to be able to clearly distinguish observable behaviors and assign ratings in a
consistent manner. This training should also include modeling and instruction for the execution of meaningful dialog and discussion during a post observation conference. To improve the overall quality and impact that the evaluation process has on teacher practice administration must improve the quality of feedback provided during performance evaluations.

**Limitations of the Study**

There were three predominate limitations contained in this study. First, as with all survey research, the capacity of the survey instrument used was limited by the opinions that it intended to measure. Second, limitations existed in the process for data collection. Distribution of surveys occurred during faculty meetings where participant attention was limited by fatigue, distraction of colleagues, food, and expectations of family and school obligations. These items likely influenced participants’ concentration and dedication to response.

The findings of this research are further limited due to the fact that evidence is based on County X, and may not therefore be generalizable to other school districts and other subjects because of the special focus on high school math teachers in County X.

**Recommendations for Further Research**

As previously stated, this study has limitations which confine the conclusions that can be depicted from the data gathered. Therefore, there can be several recommendations made to expand the current knowledge base with further research.

It is recommended that the range of this investigation be expanded beyond the focus of high school mathematics teachers in County X. The study population could be expanded to include math, science, English and history SOL teachers. This investigation
would further outline the relationship between teacher performance and student achievement across subject areas. This study could also be enlarged to include elementary and middle school teachers. Enlargement of this study population would give an added dimension to the perception of the quality and impact of the evaluation process has across grade levels. This added dimension would allow County X to generate best practices and needed improvements to the evaluation system for county wide revision.

It is also recommended that further investigation include the opinion of administration into the research design. Administration could give valuable input and explanation into the assignment of teacher evaluation ratings. Administrative perspective could also prove valuable in explaining the shortfall of meaningful and relevant feedback that was found in this investigation. Administrative opinion could add a dimension to this study that would not only enrich the data but could be used as a comparison with teacher perceptions of feedback that is missing from the evaluation process.

While this research investigation did include a survey which contained one open ended question for teacher response into County X evaluation system, this questions was only answered by 24% of the surveyed population. A recommendation for further study would be to expand upon the qualitative data to support the quantitative survey responses. A follow up focus group discussion of teachers and administrators reflecting on survey data would provide an additional view to investigate evaluation quality and impact that the evaluation process has on teacher practice.

Finally, a recommendation for further study would include an efficiency study of performance ratings for County X. This study would assess the consistency and
accuracy of performance ratings from one evaluator to another and evaluate if the performance indicators are clear descriptors of observed teacher behavior. If consistency is not observed then revamping of the performance indicators for each performance standard is needed.
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

References


The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district


The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district


Mahar, J. A., & Strobert, B. (2010). The use of 360-degree feedback compared to traditional evaluative feedback for the professional growth of teachers in k-12
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district


www.usca.edu/essays/vol11fall2004.html


www.uv.es/RELIEVE/v18n1/RELIEVEv181_leng.htm


http://teaching.about.com/od/gloss/g/Summative-Evaluation.html


The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district


The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district


Appendix A

Student Survey Summary

Teacher’s Name ___________________________ School Year __________________
Grade ____________________________ Subject ____________________________

1. How many surveys did you distribute?

2. How many completed surveys were returned?

3. What is the percentage of completed questionnaires you received (#1 divided into #2)?
   ____________%

Student Satisfaction Analysis

4. Describe your survey population(s) (i.e., list appropriate demographic characteristics such as grade level and subject for students).

5. List factors that might have influenced the results (e.g., survey was conducted as the bell rang for dismissal).

6. Analyze survey responses and answer the following questions:
   A) What did students perceive as your major strengths?
   B) What did students perceive as your major weaknesses?
   C) How can you use this information for continuous professional growth?

The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Appendix A

6-12 Student Survey

The purpose of this survey is to allow you to give your teacher ideas about how this class might be improved.

**Directions:** DO NOT PUT YOUR NAME ON THIS SURVEY. Write your teacher’s name, school year, and class period in the space provided. Listed below are several statements about this class. Indicate your agreement with each statement. If you disagree, circle 1; if you agree circle 3; if you are undecided, circle 2. If you wish to comment, please write your comments at the end of the survey.

Teacher’s Name ______________________  School Year ______________  Class Period ______

<table>
<thead>
<tr>
<th>In this class, my teacher...</th>
<th>Disagree</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. gives clear instructions.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2. treats everyone fairly.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3. is available for help outside of class time.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4. clearly states the objectives for the lesson.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5. grades my work in a reasonable time.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6. relates the lesson to other subjects or the real world.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7. allows for and respects different opinions.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8. encourages all students to learn.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9. uses a variety of activities.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10. communicates in a way I can understand.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11. manages the classroom with a minimum of disruptions.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12. shows respect to all students.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>13. makes sure class time is used for learning.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>14. clearly defines long-term assignments(such as projects)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>15. sets high expectations.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16. helps me reach the high expectations she/he sets.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17. communicates honestly with me.</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

COMMENTS:

Appendix B

PERFORMANCE STANDARDS
Performance Standard 1: Data-Driven Planning
The teacher uses data to plan appropriate curricula, implement instructional strategies, and use resources to promote learning for all students.

Sample Performance Indicators
The teacher:
♦ Designs coherent instruction based upon knowledge of subject matter and student performance data.
♦ Plans instruction to achieve desired objectives that reflect the Virginia Standards of learning and division curriculum guides.
♦ Identifies and plans for the instructional and developmental needs of all students.
♦ Selects varied and appropriate instructional strategies and materials.
♦ Includes specific student performance expectations in instructional planning.
♦ Develops plans that address immediate and long-range goals.

Performance Rubric

<table>
<thead>
<tr>
<th>Exceeds Standard</th>
<th>Meets Standard*</th>
<th>Needs Improvement</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition to meeting the standard…</td>
<td>The teacher successfully uses data to optimize resources in the planning process.</td>
<td>The teacher uses data to plan appropriate curricula, implement instructional strategies, and use resources to promote learning for all students.</td>
<td>The teacher inconsistently uses data in the instructional planning process.</td>
</tr>
</tbody>
</table>

*“Meets Standard” is the baseline of acceptable performance for teachers and is the actual performance standard.

Documentation Log
No documentation is required as part of the Documentation Log as teachers are responsible for developing, maintaining, and adapting long- and short-term lesson plans. The lesson plans are observable through items other than the log such as lesson plan submissions or the lesson plan book.

Appendix B

Performance Standard 2: Instructional Delivery
The teacher promotes student learning by addressing individual learning differences and by using effective instructional strategies.

Sample Performance Indicators
The teacher:
♦ Incorporates a variety of teaching methods and instructional strategies in lessons.
♦ Uses comprehensive materials, technology, and resources to support student learning.
♦ Teaches essential knowledge, and develops students’ critical thinking and problem-solving skills.
♦ Makes learning relevant by connecting students’ prior knowledge and experiences to the learning process.
♦ Engages and maintains students in active learning.
♦ Differentiates instruction based on student diversity and individual needs.

Performance Rubric

<table>
<thead>
<tr>
<th>Exceeds Standard</th>
<th>Meets Standard*</th>
<th>Needs Improvement</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition to meeting the standard…</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The teacher successfully meets the individual learning needs of all student groups through effective instruction within a variety of settings.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The teacher promotes student learning by addressing individual learning differences and by using effective instructional strategies.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The teacher does not consistently address individual learning differences and/or use effective instructional strategies.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The teacher does not effectively deliver instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* “Meets Standard” is the baseline of acceptable performance for teachers and is the actual performance standard.

Documentation Log
No documentation is required as instructional delivery is the focus of classroom observation. Teachers may maintain copies of their classroom observation forms in this section.

Appendix B

Performance Standard 3: Assessment
The teacher analyzes assessment data to measure student progress and guide immediate and long-range instruction.

Sample Performance Indicators
The teacher:
♦ Uses a variety of informal and formal assessments.
♦ Provides timely and specific feedback.
♦ Collects and maintains assessment data records.
♦ Analyzes and interprets data.
♦ Uses analysis and interpretation data to guide instructional decisions (e.g., reteaches and/or accelerates).
♦ Provides self-assessment strategies for students.

Performance Rubric

<table>
<thead>
<tr>
<th>Exceeds Standard</th>
<th>Meets Standard*</th>
<th>Needs Improvement</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition to meeting the standard…</td>
<td>The teacher analyzes assessment data to measure student progress and guide immediate and long-range instruction.</td>
<td>The teacher inconsistently: uses a variety of assessment strategies, links assessment to intended learning outcomes, modifies instruction based on assessment data, and/or reports student progress in a timely fashion.</td>
<td>The teacher infrequently: conducts assessments, uses a range of assessment formats, and/or applies assessment data to the instructional decision making process.</td>
</tr>
</tbody>
</table>

*“Meets Standard” is the baseline of acceptable performance for teachers and is the actual performance standard.

Documentation Log
Check the box(es) below to indicate required documentation item(s) included.
Year 1  ✖  Grading procedures
Continuing Contract Teachers Only
Year 2  ✖  Grading procedures
Year 3  ✖  Grading procedures

Appendix B

Performance Standard 4: Learning Environment

The teacher provides a well-managed, safe student-centered environment that is academically challenging.

Sample Performance Indicators

The teacher:
♦ Establishes rapport in a climate of trust and respect.
♦ Recognizes and fosters appreciation of diversity.
♦ Engages students in the learning process.
♦ Implements classroom and school rules and routines fairly and consistently.
♦ Provides a safe and positive learning environment.
♦ Maximizes instructional time.
♦ Facilitates a student-centered learning environment.

Performance Rubric

<table>
<thead>
<tr>
<th>Exceeds Standard</th>
<th>Meets Standard*</th>
<th>Needs Improvement</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition to meeting the standard…</td>
<td>The teacher provides a well-managed, safe student-centered environment that is academically challenging.</td>
<td>The teacher inconsistently demonstrates expectations for student behavior and/or academic achievement.</td>
<td>The teacher rarely maintains acceptable expectations for student behavior and/or academic achievement</td>
</tr>
</tbody>
</table>

* “Meets Standard” is the baseline of acceptable performance for teachers and is the actual performance standard.

Documentation Log

Check the boxes below to indicate required documentation items included.

Year 1  ■ Classroom rules/discipline plan
       ■ Student survey summary

Continuing Contract Teachers Only

Year 2  ■ Classroom rules/discipline plan
        ■ Student survey summary

Year 3  ■ Classroom rules/discipline plan
        ■ Student survey summary

The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

Appendix B

Performance Standard 5: Communication
The teacher communicates effectively with students, staff, parents/guardians, and the community.

Sample Performance Indicators
The teacher:
♦ Uses precise language and acceptable forms of oral and written expression.
♦ Explains directions, concepts, and lesson content to students in a logical, sequential, and age appropriate manner.
♦ Shares major instructional goals and classroom expectations with students and parents/guardians.
♦ Initiates communication and responds to parents/guardians regarding student expectations, progress, or concerns in a timely and confidential manner.

Performance Rubric

<table>
<thead>
<tr>
<th>Exceeds Standard</th>
<th>Meets Standard*</th>
<th>Needs Improvement</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>In addition to meeting the standard…</em></td>
<td>The teacher communicates effectively with students, staff, parents/guardians, and the community.</td>
<td>The teacher is ineffective in communicating with students, staff, parents, or community and/or inconsistently communicates concepts and class expectations to students.</td>
<td>The teacher consistently fails to communicate and respond to student, staff, parent, or community concerns and/or poorly articulates content and expectations to students.</td>
</tr>
<tr>
<td>The teacher clearly communicates expectations and content to all student groups in a variety of ways and initiates communication with parents and the community.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*“Meets Standard” is the baseline of acceptable performance for teachers and is the actual performance standard.

Documentation Log
Check the boxes below to indicate required documentation items included.
Year 1
- Parent contact log
- Conference log
- Long-range plans/course syllabus

Continuing Contract Teachers Only
Year 2
- Parent contact log
- Conference log
- Long-range plans/course syllabus

Year 3
- Parent contact log
- Conference log
- Long-range plans/course syllabus

Effective Teacher Research
Appendix B

Performance Standard 6: Professionalism
The teacher maintains a professional demeanor, participates in professional growth opportunities, demonstrates an understanding of the curriculum, and contributes to the profession.

Sample Performance Indicators
The teacher:
♦ Maintains a positive pattern of professional behavior (e.g., appearance, punctuality, and attendance).
♦ Respects and maintains confidentiality.
♦ Performs assigned school duties and follows policies and procedures.
♦ Demonstrates knowledge and skills relevant to the subject area(s) taught.
♦ Evaluates and identifies areas of personal strengths and weaknesses related to professional skills and their impact on student learning.
♦ Sets goals for improvement of skills and professional performance.
♦ Participates in professional growth activities and incorporates learning into instructional practice.
♦ Serves on school and/or division committees and supports school activities.

Performance Rubric

<table>
<thead>
<tr>
<th>Exceeds Standard in addition to meeting the standard…</th>
<th>Meets Standard*</th>
<th>Needs Improvement</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>The teacher is a professional role model for others, engages in a high level of professional growth, and contributes to the development of others and the well-being of the profession.</td>
<td>The teacher maintains a professional demeanor, participates in professional growth opportunities, demonstrates an understanding of the curriculum, and contributes to the profession.</td>
<td>The teacher inconsistently: participates in professional growth activities, applies strategies and information from professional growth opportunities, serves the profession, and/or demonstrates professional judgment.</td>
<td>The teacher demonstrates inflexibility, a reluctance to support others in the work of the school, and/or rarely takes advantage of professional growth opportunities.</td>
</tr>
</tbody>
</table>

*“Meets Standard” is the baseline of acceptable performance for teachers and is the actual performance standard.

Documentation Log
Check the box(es) below to indicate required documentation item(s) included.
Year 1 Professional development log
Continuing Contract Teachers Only
Year 2 Professional development log
Year 3 Professional development log

Appendix B

Performance Standard 7: Student Achievement
The work of the teacher results in acceptable, measurable student progress.

Sample Performance Indicators
The teacher:
♦ Sets measurable and appropriate achievement goals for student progress.
♦ Uses assessment data to regularly monitor student progress and modify instruction as needed.
♦ Identifies and establishes additional means of support to increase the achievement level for all groups of students.
♦ Provides evidence that achievement goals have been met.

Performance Rubric

<table>
<thead>
<tr>
<th>Exceeds Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition to meeting the standard…</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meets Standard*</th>
</tr>
</thead>
<tbody>
<tr>
<td>The work of the teacher results in acceptable, measurable student progress.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>The work of the teacher results in an acceptable level of achievement for some subgroups.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>The work of the teacher does not result in an acceptable level of achievement for most subgroups.</td>
</tr>
</tbody>
</table>

*“Meets Standard” is the baseline of acceptable performance for teachers and is the actual performance standard.

Documentation Log

Check the boxes below to indicate required documentation items included.

Year 1
- Academic goal-setting form(s)
- Documentation of student progress relating to the goal(s) set

Continuing Contract Teachers Only

Year 2
- Academic goal-setting form(s)
- Documentation of student progress relating to the goal(s) set

Year 3
- Academic goal-setting form(s)
- Documentation of student progress relating to the goal(s) set

Appendix C

Survey

Sections 1: Demographic Information

1. How many years, including the current year, have you taught for Bedford County Public Schools
   a. 1-3 years
   b. 4-7 years
   c. 8-11 years
   d. 12-19 years
   e. 20+ years
2. How many years, including the current year, have you taught.
   a. 1-3 years
   b. 4-7 years
   c. 8-11 years
   d. 12-19 years
   e. 20+ years
3. Choose the department that you are most closely associated with in your current teaching assignment.
   a. Mathematics
   b. Science
   c. English
   d. Social Studies
   e. Foreign Language
   f. Related Arts
   g. Health PE
   h. Special Education
4. Your gender
   a. Male
   b. Female
5. Your age
   a. 21-30
   b. 31-40
   c. 41-50
   d. 51-60
   e. 61+
6. Level of Education
   a. Bachelor Degree
   b. Master’s Degree
   c. Doctorate Degree
7. The score you received on the majority of the performance standards on your last summative evaluation (4 out of 7).
   a. Exceeds Standard
   b. Meets Standard
   c. Needs Improvement
   d. Unsatisfactory
   e. Do not know score

Sections 2: Validity

Please reflect on your last summative performance evaluation and answer the following questions about the Bedford County Public School Teacher Performance Evaluation.

8. The results of my summative evaluation accurately portrayed my performance
   
   Strongly Disagree 1 2 3 4 5 Strongly Agree

9. Clear examples were given to justify the ratings I received on my summative evaluation.

   Strongly Disagree 1 2 3 4 5 Strongly Agree

10. The rating I received on my summative evaluation did not reflect bias from my evaluator.

   Strongly Disagree 1 2 3 4 5 Strongly Agree

11. The number of observations and times spent conducting classroom observations including pre and post observation conferences allows for an accurate prediction of my teaching performance.

   Strongly Disagree 1 2 3 4 5 Strongly Agree

12. The measurement of student performance that is evaluated during the goal setting process provided an accurate picture of my ability to impact student learning.

   Strongly Disagree 1 2 3 4 5 Strongly Agree

13. The information collected in the documentation log provides evidence of several performance standards. These items are a valid way of providing a comprehensive portrait of my work.

   Strongly Disagree 1 2 3 4 5 Strongly Agree

14. The teacher evaluation instrument used to rate my performance is fair and valid.

   Strongly Disagree 1 2 3 4 5 Strongly Agree
Section 3: Evaluation Feedback

Please reflect on the feedback you received during your last post observation conference to answer the following questions.

15. Amount of information received
   None 1 2 3 4 5 Great Deal

16. Depth of information received
   Shallow 1 2 3 4 5 In-Depth

17. Quality of the ideas and suggestions contained in the feedback
   Low 1 2 3 4 5 High

18. Specificity of information provided
   General 1 2 3 4 5 Specific

19. Timing of the feedback
   Delayed 1 2 3 4 5 Immediate

20. Application of information toward student achievement
   Low 1 2 3 4 5 High

21. Feedback was useful for my professional development
   Strongly Disagree 1 2 3 4 5 Strongly Agree

22. Feedback was meaningful and assisted me to improve my classroom instruction
   Strongly Disagree 1 2 3 4 5 Strongly Agree

Section 4: Multiple Sources of Data Used

Please reflect over the last year and report the extent the following sources of performance information were considered as part of formative and summative evaluations.

23. Observation of your classroom performance
   Not Considered 1 2 3 4 5 Used extensively

24. Examination of artifacts in your documentation log
   Not Considered 1 2 3 4 5 Used extensively
25. Student Survey Summary

Not Considered  1  2  3  4  5  Used extensively

26. Goal setting to assess student achievement

Not Considered  1  2  3  4  5  Used extensively

**Section 5: Reflection on Practice**

Please look back over the last year and report the extent the following sources of performance information caused you to reflect and improve your teaching practice.

27. Feedback from evaluator given during post observation conference

Little Reflection  1  2  3  4  5  Great Reflection

28. Collection of artifacts in your documentation log

Little Reflection  1  2  3  4  5  Great Reflection

29. Completion of the student survey and student survey summary

Little Reflection  1  2  3  4  5  Great Reflection

30. Data analysis to assess student achievement as part of the goal setting process

Little Reflection  1  2  3  4  5  Great Reflection

**Section 6: Overall Rating**

Please reflect on your most recent summative evaluation experience in Bedford County. Consider the entire evaluation process including observations, goal setting, documentation log, student survey summary, feedback, etc.

31. Rate the overall quality of the evaluation

Very Poor Quality  1  2  3  4  5  Very High Quality

32. Rate the overall impact of the evaluation on your teaching practices. (A strong impact, rating of 5, would imply profound changes in your teaching practices while low impact, a rating of 1, would imply no changes in your teaching practices).

No Impact  1  2  3  4  5  Strong Impact
33. Please write any comments you may have about the Bedford County Teacher Performance Evaluation System below.
Appendix D

**Holistic Performance Assessment Score**

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Holistic Assessment Admin 1</th>
<th>Holistic Assessment Admin 2</th>
<th>Holistic Assessment Admin 3</th>
<th>Holistic Assessment Admin 4</th>
<th>(Mean) Holistic Performance Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS1T1</td>
<td>2.5</td>
<td>3.0</td>
<td>3.0</td>
<td>3</td>
<td>2.875</td>
</tr>
<tr>
<td>HS1T2</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3</td>
<td>3.000</td>
</tr>
<tr>
<td>HS1T3</td>
<td>2.5</td>
<td>2.5</td>
<td>2.0</td>
<td>3</td>
<td>2.375</td>
</tr>
<tr>
<td>HS1T4</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3</td>
<td>3.000</td>
</tr>
<tr>
<td>HS1T5</td>
<td>2.0</td>
<td>2.5</td>
<td>2.0</td>
<td>2</td>
<td>2.125</td>
</tr>
<tr>
<td>HS1T6</td>
<td>3.5</td>
<td>4.0</td>
<td>3.5</td>
<td>4</td>
<td>3.750</td>
</tr>
<tr>
<td>HS1T7</td>
<td>3.0</td>
<td></td>
<td>3.0</td>
<td>3</td>
<td>3.000</td>
</tr>
<tr>
<td>HS1T8</td>
<td>2.0</td>
<td>2.0</td>
<td>2.5</td>
<td>2</td>
<td>2.125</td>
</tr>
<tr>
<td>HS1T9</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4</td>
<td>4.000</td>
</tr>
<tr>
<td>HS1T10</td>
<td>3.5</td>
<td>3.0</td>
<td>3.0</td>
<td>3</td>
<td>3.125</td>
</tr>
<tr>
<td>HS1T11</td>
<td>4.0</td>
<td>3.5</td>
<td>3.5</td>
<td>4</td>
<td>3.750</td>
</tr>
<tr>
<td>HS2T1</td>
<td>Not observed</td>
<td>3.0</td>
<td>Not Observed</td>
<td></td>
<td>3.000</td>
</tr>
<tr>
<td>HS2T2</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td></td>
<td>4.000</td>
</tr>
<tr>
<td>HS2T3</td>
<td>Not observed</td>
<td>3.0</td>
<td>Not Observed</td>
<td></td>
<td>3.000</td>
</tr>
<tr>
<td>HS2T4</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td>3.000</td>
</tr>
<tr>
<td>HS2T5</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td>3.000</td>
</tr>
<tr>
<td>HS2T6</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td>3.000</td>
</tr>
<tr>
<td>HS2T7</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td></td>
<td>1.330</td>
</tr>
<tr>
<td>HS2T8</td>
<td>2.0</td>
<td>2.5</td>
<td>3.0</td>
<td></td>
<td>2.500</td>
</tr>
<tr>
<td>HS3T1</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td>3.000</td>
</tr>
<tr>
<td>HS3T2</td>
<td>3.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
<td>2.330</td>
</tr>
<tr>
<td>HS3T3</td>
<td>3.0</td>
<td>3.0</td>
<td>1.0</td>
<td></td>
<td>2.330</td>
</tr>
<tr>
<td>HS3T4</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
<td>2.000</td>
</tr>
<tr>
<td>HS3T5</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td>3.000</td>
</tr>
<tr>
<td>HS3T6</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td>3.000</td>
</tr>
</tbody>
</table>

NOTE: High School 1 (HS1) has four administrators. High School 2 and 3 (HS2, HS3) only have three administrators. The holistic performance assessment score is obtained by averaging scores of each administrator that had directly observed a teacher’s performance.
Appendix E

Student Performance Score

<table>
<thead>
<tr>
<th>Teacher</th>
<th>SOL Subject</th>
<th>Failed</th>
<th>Passed Proficient</th>
<th>Passed Advanced</th>
<th>Student Performance Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS1T1</td>
<td>Algebra II</td>
<td>49</td>
<td>21</td>
<td>2</td>
<td>382.56</td>
</tr>
<tr>
<td>HS1T2</td>
<td>Geometry</td>
<td>22</td>
<td>88</td>
<td>13</td>
<td>438.03</td>
</tr>
<tr>
<td>HS1T3</td>
<td>Algebra I</td>
<td>15</td>
<td>16</td>
<td>0</td>
<td>395.16</td>
</tr>
<tr>
<td>HS1T4*</td>
<td>Algebra II</td>
<td>0</td>
<td>17</td>
<td>6</td>
<td>470.43</td>
</tr>
<tr>
<td>HS1T4*</td>
<td>Geometry</td>
<td>45</td>
<td>45</td>
<td>0</td>
<td>389.45</td>
</tr>
<tr>
<td>HS1T5</td>
<td>Geometry</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>402.92</td>
</tr>
<tr>
<td>HS1T6</td>
<td>Algebra I</td>
<td>28</td>
<td>20</td>
<td>0</td>
<td>386.83</td>
</tr>
<tr>
<td>HS1T7</td>
<td>Algebra II</td>
<td>60</td>
<td>42</td>
<td>1</td>
<td>371.38</td>
</tr>
<tr>
<td>HS1T8</td>
<td>Algebra I</td>
<td>36</td>
<td>17</td>
<td>0</td>
<td>379.75</td>
</tr>
<tr>
<td>HS1T9</td>
<td>Algebra II</td>
<td>21</td>
<td>68</td>
<td>34</td>
<td>436.99</td>
</tr>
<tr>
<td>HS1T10</td>
<td>Geometry</td>
<td>0</td>
<td>58</td>
<td>19</td>
<td>473.01</td>
</tr>
<tr>
<td>HS1T11</td>
<td>Algebra II</td>
<td>0</td>
<td>22</td>
<td>15</td>
<td>497.62</td>
</tr>
<tr>
<td>HS2T1</td>
<td>Algebra II</td>
<td>78</td>
<td>27</td>
<td>0</td>
<td>370.90</td>
</tr>
<tr>
<td>HS2T2</td>
<td>Geometry</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>307.00</td>
</tr>
<tr>
<td>HS2T3</td>
<td>Geometry</td>
<td>37</td>
<td>11</td>
<td>0</td>
<td>372.19</td>
</tr>
<tr>
<td>HS2T4</td>
<td>Algebra I</td>
<td>46</td>
<td>16</td>
<td>0</td>
<td>379.79</td>
</tr>
<tr>
<td>HS2T5</td>
<td>Geometry</td>
<td>6</td>
<td>36</td>
<td>1</td>
<td>425.72</td>
</tr>
<tr>
<td>HS2T6</td>
<td>Algebra II</td>
<td>0</td>
<td>22</td>
<td>2</td>
<td>443.95</td>
</tr>
<tr>
<td>HS2T7</td>
<td>Algebra I</td>
<td>67</td>
<td>10</td>
<td>0</td>
<td>369.92</td>
</tr>
<tr>
<td>HS2T8</td>
<td>Algebra II</td>
<td>101</td>
<td>7</td>
<td>0</td>
<td>344.16</td>
</tr>
<tr>
<td>HS3T1*</td>
<td>Algebra I</td>
<td>54</td>
<td>43</td>
<td>0</td>
<td>372.47</td>
</tr>
<tr>
<td>HS3T1*</td>
<td>Geometry</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>413.35</td>
</tr>
<tr>
<td>HS3T2</td>
<td>Geometry</td>
<td>32</td>
<td>23</td>
<td>1</td>
<td>378.95</td>
</tr>
<tr>
<td>HS3T3*</td>
<td>Alg I</td>
<td>14</td>
<td>8</td>
<td>0</td>
<td>384.77</td>
</tr>
<tr>
<td>HS3T3*</td>
<td>Geometry</td>
<td>50</td>
<td>32</td>
<td>0</td>
<td>387.10</td>
</tr>
<tr>
<td>LHST4</td>
<td>Algebra II</td>
<td>58</td>
<td>8</td>
<td>0</td>
<td>349.66</td>
</tr>
<tr>
<td>LHST5*</td>
<td>Algebra II</td>
<td>36</td>
<td>9</td>
<td>0</td>
<td>377.73</td>
</tr>
<tr>
<td>LHST5*</td>
<td>Algebra I</td>
<td>34</td>
<td>5</td>
<td>0</td>
<td>371.03</td>
</tr>
<tr>
<td>LHST6</td>
<td>Algebra II</td>
<td>8</td>
<td>43</td>
<td>3</td>
<td>469.33</td>
</tr>
</tbody>
</table>

NOTE: The student performance score was calculated by each teacher by adding up all the SOL scores per subject area (Algebra I, Geometry, Algebra II) and dividing by the total number of scores. This average of the SOL scores was reported as the student performance score. * Indicated that the teacher taught two different SOL subjects and therefore had two student performance scores.
Appendix F

Survey Questions Mapped to Indices and Descriptive Statistics

Section 2: Validity (Cronback alpha = .883; mean = 3.69; std = .860; min =1.0; max =5.0) For each of the below statements, please reflect on your last summative performance evaluation and answer the following questions about County X Teacher Performance Evaluation by indicating the extent to which you agree or disagree (1 strongly disagree to 5 strongly agree).

- The results of my summative evaluation accurately portrayed my performance
- Clear examples were given to justify the ratings I received on my summative evaluation.
- The rating I received on my summative evaluation did not reflect bias from my evaluator.
- The number of observations and times spent conducting classroom observations including pre and post observation conferences allows for an accurate prediction of my teaching performance.
- The measurement of student performance that is evaluated during the goal setting process provided an accurate picture of my ability to impact student learning.
- The information collected in the documentation log provides evidence of several performance standards. These items are a valid way of providing a comprehensive portrait of my work.
- The teacher evaluation instrument used to rate my performance is fair and valid.

Section 3: Evaluation Feedback (Cronbach alpha = .935; mean= 3.61; std=.838; min=1.13; max=5) For each of the questions listed below please reflect on the feedback you received during you last post observation conference and respond with the corresponding scale provided.

- Amount of information received
  None 1 2 3 4 5 Great Deal
- Depth of information received
  Shallow 1 2 3 4 5 In-Depth
- Quality of the ideas and suggestions contained in the feedback
  Low 1 2 3 4 5 High
- Specificity of information provided
  General 1 2 3 4 5 Specific
- Timing of the feedback
  Delayed 1 2 3 4 5 Immediate
- Application of information toward student achievement
  Low 1 2 3 4 5 High
Appendix F (continued)

Survey Questions Mapped to Indices and Descriptive Statistics

- Feedback was useful for my professional development
  Strongly Disagree 1 2 3 4 5 Strongly Agree
- Feedback was meaningful and assisted me to improve my classroom instruction
  Strongly Disagree 1 2 3 4 5 Strongly Agree

Section 4: Multiple Sources of Data Used (Cronbach alpha = .725; mean = 3.29; std = .824; min= 1.25; max= 5.0) For each of the below statements, please reflect over the last year and report the extent the following sources of performance information were considered as part of formative and summative evaluations (1 not considered to 5 used extensively).

- Observation of your classroom performance
- Examination of artifacts in your documentation log
- Student Survey Summary
- Goal setting to assess student achievement

Section 5: Reflection on Practice (Cronbach alpha = .770; mean=3.61; std=.838; min= 1.0; max=5.0) For each of the below statements, please reflect over the last year and report the extent the following sources of performance information caused you to reflect and improve your teaching practice (1 little reflection to 5 great reflection).

- Feedback from evaluator given during post observation conference
- Collection of artifacts in your documentation log
- Completion of the student survey and student survey summary
- Data analysis to assess student achievement as part of the goal setting process
Date: May 6, 2013
To: Lee Ann Calvert
Re: Approval of Research Proposal

Your request for an expedited review of your research project: “Do Teacher Performance Evaluations Impact Teacher Instructional Practices & Predict Student Achievement? Evidence from a Rural High School” has been completed. The proposal and related study comply with the standards set by the U.S. Department of Health and Human Services, Code of Federal Regulations, Title 45 CFR Part 46, Protection of Human Subjects, effective as of July 14, 2009. The study is therefore approved.

Please remember that if any modifications are necessary, these changes need to be approved by this committee. Approval for this proposal is for one year. If necessary, re-approval must occur prior to May 5, 2014. Please feel free to give me a call at X8962 if you have any questions.

Sincerely,

Beth McKinney

Beth McKinney, PhD, MPH, CHES
Chair, Human Subject Research Committee (IRB)
The impact of performance evaluations on teachers’ instructional practices and student achievement. Evidence from a rural school district

---

1 Hard to staff schools are those schools that consistently have trouble finding and retaining qualified teachers. Schools which are located in economically depressed areas or schools in isolated districts that do not have funding or amenities available to find and retain qualified teachers are considered hard to staff. These school districts experience high turnover rates because once teachers achieve experience they leave for higher salaries and better working conditions (NEA, n.d.).

2 President Obama’s Race to the Top (RTTT) program, backed by 4.35 billion dollars in federal funding, awards money to states that develop and implement rigorous standards and quality assessments that measure student knowledge and growth. The goal of this program is to prepare all students for success in college or the workforce and to restore the nation as a leader in college graduates. The expectation of this program includes improvement of teacher preparation and revision of teacher evaluation and compensation with the purpose of teachers being rewarded based on effectiveness (The White House, 2009).

3 A performance standard is defined as the major duties and responsibilities performed by a teacher (Carnot, G. et al., 2007).