2013

Teacher perspectives of the use of student performance data in teacher evaluations

Paul Thomas Hopkins
College of William & Mary - School of Education

Follow this and additional works at: https://scholarworks.wm.edu/etd

Part of the Educational Administration and Supervision Commons, Educational Assessment, Evaluation, and Research Commons, and the Educational Leadership Commons

Recommended Citation
https://dx.doi.org/doi:10.25774/w4-26gm-ny79

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
TEACHER PERSPECTIVES OF THE USE OF
STUDENT PERFORMANCE DATA IN TEACHER EVALUATIONS

A Dissertation

Presented to

The Faculty of the School of Education

The College of William and Mary in Virginia

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Education

By
Paul Thomas Hopkins
September 2013
Table of Contents

Acknowledgements v
List of Tables vi
List of Charts viii
Abstract ix

Chapter 1: The Problem
  Statement of Purpose 6
  Research Questions 6
  Significance of the Study 7
  Justification 9
  Operational Definitions of Key Terms 10
  Delimitations and Limitations 12

Chapter 2: Relevant Literature
  Historical Overview of Teacher Evaluation Programs 15
  Teacher Perceptions to Teacher Evaluation Methods 25
  Student Performance Data in Education Accountability Programs 41
  Introduction of Student Performance Data to Teacher Evaluations 55
  Summary 62

Chapter 3: Methodology
  Research Questions 66
  Sample 66
  Instrumentation 67
  Procedures 75
  Data Analysis 76
  Ethical Considerations 77

Chapter 4: Analysis of Results
  The Study 81
  Findings for Research Question One 86
  Findings for Research Question Two 91
  Findings for Research Question Three 96
  Findings for Research Question Four 100
  Findings for Research Question Five 105

Chapter 5: Summary and Discussion of Findings
  Discussion of Findings Related to Research Questions 116
  General Discussion of Findings 131
  Conclusion 141
  Delimitations and Limitations 142
  Recommendations for Future Research 143
Acknowledgements

The successful completion of this journey is a direct result of many individuals. The experience and insight of many professors at The College of William & Mary made this possible. Dr. James Stronge, the chairman of my dissertation committee and my academic advisor, provided me with ongoing support and guidance. His expertise, insight, and encouragement were invaluable to my success throughout the dissertation study process. Dr. Christopher Gareis and Dr. Tom Ward were also extremely supportive and provided valuable guidance and constructive feedback during the development, writing, and data analysis of this project. Dr. Steven Staples was also instrumental in the development of this study and was a positive influence and mentor throughout my course work and professional career. In addition to these esteemed professors, I am extremely grateful to Dr. Jennifer Parish, Dr. Megan Tschannen-Moran, Dr. Michael DiPaola, and many others for their guidance and mentorship during my graduate and doctoral course work. I feel privileged to have completed my graduate and doctoral program at such a respected and collegial academic community.

I am extremely indebted to a number of individuals who provided me with unwavering support throughout my doctoral studies. My wife, Megan, and two children, Madalyn and Jarrett, were always understanding of the time I was away in class or at the library completing research. Their unconditional love and support made the completion of this study a reality. My parents, Norman and Judy Hopkins, siblings, and extended family were also influential in providing me the steadfast support to embark upon and finish this journey. I am truly blessed to have such a committed family.

Finally, the completion of this project was also assisted by a number of colleagues at The College of William & Mary and in my professional circles. Dr. Rob Williams, Dr. Lisa Pennycuff, Amy Colley, and April Lawrence were grateful with their time, expertise, and support during this project. I am honored to have learned from a number of personal and professional colleagues and friends who have helped individually and collectively make me a more effective instructional leader.
List of Tables

2. Summary of Evaluation Models 39
3. Description of Utility Standards 50
4. Description of Feasibility Standards 51
5. Description of Propriety Standards 52
6. Description of Accuracy Standards 53
7. Panel of Experts’ Feedback on Survey 70
8. Table of Specifications for Survey 72
9. Data Analysis Table 76
10. Response Rates following Each Email Deployment 82
11. Homogeneity of Responses 82
12. Participants’ Years of Experience 83
13. Participants in a Collective Bargaining Agreement (CBA) 84
14. Participants in a Tested or Non-Tested Grade or Course 85
15. Participants’ Level of School 85
16. Experience with an Evaluation Program that Utilizes Student Performance Data (SPD) 86
17. Propriety Standard Descriptive Statistics 87
18. Cronbach’s Alpha for Propriety Standard 88
19. Mean and Standard Deviation for Propriety Standard Questions By Participant’s Experience with an Evaluation Program that Utilizes Student Performance Data (SPD) 89
20. Propriety Standard ANOVA by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data 89
21. Tukey Post-hoc Analysis for Propriety Standard by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data 90
22. Teacher-cited Benefits and Limitations Associated with the Propriety Standard 91
23. Utility Standard Descriptive Statistics 92
24. Cronbach’s Alpha for Utility Standard 93
25. Mean and Standard Deviation for Utility Standard Questions By Participant’s Experience with an Evaluation Program that Utilizes Student Performance Data (SPD) 94
26. Utility Standard ANOVA by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data 95
27. Tukey Post-hoc Analysis for Utility Standard by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data 95
28. Teacher-cited Benefits and Limitations Associated with the Utility Standard 96
29. Feasibility Standard Descriptive Statistics 97
30. Cronbach’s Alpha for Feasibility Standard 97
31. Mean and Standard Deviation for Feasibility Standard Questions
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Participant’s Experience with an Evaluation Program that</td>
<td></td>
</tr>
<tr>
<td>Utilizes Student Performance Data (SPD)</td>
<td>98</td>
</tr>
<tr>
<td>32. Feasibility Standard ANOVA by Teacher Experience with an</td>
<td>99</td>
</tr>
<tr>
<td>Evaluation Program that Utilizes Student Performance Data</td>
<td></td>
</tr>
<tr>
<td>33. Teacher-cited Benefits and Limitations Associated with the</td>
<td>99</td>
</tr>
<tr>
<td>Feasibility Standard</td>
<td></td>
</tr>
<tr>
<td>34. Accuracy Standard Descriptive Statistics</td>
<td>101</td>
</tr>
<tr>
<td>35. Cronbach’s Alpha for Accuracy Standard</td>
<td>102</td>
</tr>
<tr>
<td>36. Mean and Standard Deviation for Accuracy Standard Questions</td>
<td>103</td>
</tr>
<tr>
<td>By Participant’s Experience with an Evaluation Program that</td>
<td></td>
</tr>
<tr>
<td>Utilizes Student Performance Data (SPD)</td>
<td></td>
</tr>
<tr>
<td>37. Accuracy Standard ANOVA by Teacher Experience with an</td>
<td>104</td>
</tr>
<tr>
<td>Evaluation Program that Utilizes Student Performance Data</td>
<td></td>
</tr>
<tr>
<td>38. Tukey Post-hoc Analysis for Accuracy Standard by Teacher</td>
<td>104</td>
</tr>
<tr>
<td>Experience with an Evaluation Program that Utilizes Student Performance Data</td>
<td></td>
</tr>
<tr>
<td>39. Teacher-cited Benefits and Limitations Associated with the</td>
<td>105</td>
</tr>
<tr>
<td>Accuracy Standard</td>
<td></td>
</tr>
<tr>
<td>40. Teacher ANOVA by Years of Experience</td>
<td>106</td>
</tr>
<tr>
<td>41. Teacher ANOVA by Participation under a CBA</td>
<td>107</td>
</tr>
<tr>
<td>42. Teacher ANOVA by Tested and Non-Tested Grades and Courses</td>
<td>108</td>
</tr>
<tr>
<td>43. Teacher ANOVA by Level of School</td>
<td>109</td>
</tr>
<tr>
<td>44. Teacher Mean &amp; Standard Deviation (SD) for Evaluation Standards</td>
<td>110</td>
</tr>
<tr>
<td>Disaggregated by Teacher Experience with an</td>
<td></td>
</tr>
<tr>
<td>Evaluation Program that Utilizes Student Performance Data</td>
<td></td>
</tr>
<tr>
<td>45. Cronbach’s Alpha Analyses for Four Evaluation Standards</td>
<td>133</td>
</tr>
</tbody>
</table>
List of Charts

1. Teacher Mean for Evaluation Standards Disaggregated by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data .............................. 111
2. Responses to Propriety Questions by Participants' Experience with Students Performance Data (SPD) in Evaluation Program ...................... 119
3. Responses to Utility Questions by Participants' Experience with Student Performance Data (SPD) in Evaluation Program .......................... 123
4. Responses to Accuracy Questions by Participants' Experience with Student Performance Data (SPD) in Evaluation Program ...................... 127
TEACHER PERSPECTIVES OF THE USE OF
STUDENT PERFORMANCE DATA IN TEACHER EVALUATIONS

ABSTRACT

The purpose of this study was to determine how K-12 public school teachers perceive the use of student performance data in teacher evaluations. The proprietary, utility, feasibility, and accuracy standards created by the Joint Committee on Standards for Education Evaluation (JCSEE) served as a framework for the study. An online survey was deployed to a random stratified sample of teachers across the United States. Participants responded to thirty statements using a four-point Likert Scale that ranged from Strongly Agree to Strongly Disagree. Participants were also provided an opportunity to list and describe additional items that they favored and feared with respect to using student performance data in teacher evaluations. Descriptive statistics and ANOVA were used to determine the level of teacher agreement on statements. Responses were further analyzed to determine the impact that the following demographic factors had on perceptions: (a) years of teaching experience, (b) teaching in a union and non-union states, (c) teaching a tested and non-tested grades and courses, and (d) teaching in elementary, middle, and high schools. Results indicated a strong level of agreement among teachers on the positive impact that the use of student performance data will have on improving teacher evaluations. Further results indicated that demographic factors played a minimal role in influencing participants’ perceptions.

PAUL THOMAS HOPKINS

PROGRAM IN EDUCATIONAL POLICY, PLANNING, AND LEADERSHIP

THE COLLEGE OF WILLIAM AND MARY IN VIRGINIA
TEACHER PERSPECTIVES OF THE USE OF STUDENT PERFORMANCE DATA IN TEACHER EVALUATIONS
Chapter 1: The Problem

Teacher evaluations have experienced an unprecedented level of analysis and scrutiny in recent years. The impetus for this heightened level of oversight has been, in large part, due to a new era of accountability ushered in by the No Child Left Behind (NCLB) Act of 2001 and subsequent legislation. In 2010, the United States Department of Education specifically called for improving teacher effectiveness to ensure that every classroom has a great teacher in it (United States Department of Education, 2010). As one of four specific areas of focus, President Obama’s administration took a significant step forward in identifying teacher effectiveness as a cornerstone in its blueprint for reform in the reauthorization of the Elementary and Secondary Education Act. This was evident by the U.S. Department of Education Race to the Top Initiative which required states to take into account student growth in the design and implementation of new teacher evaluation systems (United States Department of Education, 2010).

The NCLB legislation and Obama Administration “Race to the Top” initiatives built upon the initial mandate for educational reform cited in the 1983 publication of A Nation at Risk. This report sought to raise the level of student achievement in all of America’s public schools and sparked a wave of educational reform. The report concluded that teacher evaluation systems utilized across the United States were fundamentally flawed. Teacher evaluations were specifically criticized for including performance standards and criteria that were excessively biased, subjective, and arbitrary (National Commission on Excellence in Education, 1983). The Joint Committee on Standards for Educational Evaluation (JCSEE) constructed a refined set of personnel evaluation standards to help respond to this finding. This committee created and has since expanded a set of standards that address four accepted attributes of educational evaluation: propriety, utility, feasibility, and accuracy (2009). Subsequent research affirmed that these four
attributes cited by the JCSEE are required tenants in any sound teacher evaluation (Howard & Gullickson, 2010).

The federal government’s Race to the Top initiative prompted policymakers at the national, state and local levels to develop stronger evaluation programs that more accurately identify effective teachers and, subsequently, improve student achievement. A wealth of research demonstrates that the single most important factor in a student’s level of academic achievement is predicated by the effectiveness of the student’s teacher (Aaronson et al., 2007; Goldhaber & Hansen, 2008; Hattie, 2009; Heck, 2009; Marzano, 2003a; Nye, Konstantopolulos, & Hedges, 2004; Palardy & Rumberger, 2008; Rivkin et al., 2005; Rothstein, 2010; Sass, 2008; Stronge, 2010; Stronge, Ward, & Grant, 2011). Research clearly documents the strong correlation an effective teacher has on a student’s achievement gains during the school year as measured by value-added or other test-based growth measures (Goe et al., 2008; Wright, Horn, & Sanders, 1997). Sanders and Rivers (1996) also noted that teacher effects on student academic gains can be seen as both cumulative and residual. “The core of education is teaching and learning, and the teaching-learning connection works best when we have effective teachers working with every student every day” (Stronge, 2006, p.1). Since the research clearly demonstrates that the quality of teaching matters, it is reasonable to presuppose that a quality teacher evaluation process also matters in order to know if the school system possesses high quality teachers (Stronge & Tucker, 2003). Teacher evaluations, therefore, need to acknowledge the student achievement data. State policymakers and instructional leaders must similarly acknowledge the perceptions of teachers to the use of student performance data in their evaluations.
Driven by research, government mandates and mounting public pressure, a number of states are designing and implementing teacher evaluations to align with these specifically prescribed performance standards. School districts are developing evaluation matrices that attempt to leverage the power of student performance data to complement other teacher performance domains to more effectively evaluate teachers. Virginia is joining at least 23 other states and the District of Columbia that include among others New York, North Carolina, Maryland, Ohio, Michigan and Tennessee that use student performance data to some degree in teacher evaluations (National Council on Teacher Quality, 2011). The degree to which student performance data is factored into the evaluation instrument ranges from 33-60% (New York State Education Department, n.d.). Virginia is recommending that student performance data account for 40% of a teacher's evaluation (Virginia Department of Education, 2010).

States are working to adhere to President Obama's proposal that calls for a collaborative approach to develop fair and meaningful teacher evaluation systems (United States Department of Education, 2010). While state leaders were crafting research-based evaluation instruments that now included student performance data, there was limited evidence of how teachers would perceive this change. Since teacher involvement and buy-in is critical to the lasting success of any educational reform program (Bryk et al., 1999; Greene & Lee, 2006; Schneider & Bryk, 2000; Turnbull, 2002), a study of how teachers will embrace this change is necessary.

The problem that currently exists is that there is insufficient research to discover if teachers view student performance data as a constructive component in evaluating effective instruction. In order for states and school districts to realize the expected goals from adding student performance data to teacher evaluations, it is imperative for instructional leaders to understand how teachers perceive this change. Although teacher buy-in or support is not
required, there is research that documents that educational reform programs with teacher support have greater opportunities for lasting success (Bryk et al., 1999; Greene & Lee, 2006; Schneider & Bryk, 2000; Turnbull, 2002). Once teachers understand how student performance data has the opportunity to complement other performance domains in their evaluation, the enhanced evaluation program may gain sustained support from all stakeholders. There is ample evidence that attributes teacher buy-in and support as a factor in an educational reform’s success in meeting its intended outcomes (Apaza, 2009; Clarke, 2012; Gigante & Firestone, 2008; Greer, 2006; Lasseter, 2007; Olszewski, 2009). Conversely, there is evidence where the absence of teacher buy-in crippled a program’s chance for success (Greene & Lee, 2006; Hasson, 2011; Hill 2005; Hill 2009; Orme, 2009). Teacher buy-in is a significant factor to consider and understand. Unfortunately, research on the teachers’ perception of the use of student performance data in their evaluations has been absent up to this point.

Recognizing and responding to teachers’ perceptions of the use of student performance data in evaluations has the potential to be a powerful conduit of change. Principals, for example, are discovering that student performance data serves as a valuable tool in their development as instructional leaders. The principals’ willingness to fully embrace the power of student performance data represented a key factor in the success of these new evaluation instruments. The conclusions from a study on a new evaluation program documented that both principals and the principals’ evaluators purport a high degree of satisfaction for the new evaluation instruments that include student performance data (Reid, 2006). A similar study in Pennsylvania reinforced previous research that principals who regularly review and use student performance data positively influence student achievement in schools (Soslau, 2009). This study reported that the professional practice of utilizing formative assessment data in math accounted for up to 10% of
the variability in math performance. A final study documented how principals of feeder schools effectively use student performance data to positively impact student growth and enhance vertical articulation. In collaborative sessions where value-added student data were examined, principals were able to identify elements of both effective and ineffective school programs impacting student growth (Kelsey, 2009). Again, principal buy-in and support of the use of student performance data represented the key to unlocking the power of this initiative (Reid, 2006; Soslau 2009). Teachers may discover similar results from the use of student performance data if similar buy-in and support is present or can be developed with the assistance of teacher perception studies. Data authors a compelling story for where and how reforms can enhance teaching and learning. The story can only be truly realized if stakeholders embrace and support the process.

All teachers deserve the opportunity to be evaluated utilizing objective data. Maintaining conventional evaluation programs that do not factor in student achievement data for any teacher jeopardizes opportunities for growth for teachers as well as students. Teacher evaluations have the ability to spotlight strengths in a teacher’s delivery of instruction and identify where improvement may be necessary. Although the use of student performance data is a relatively new component in teacher evaluations, it is not a novel concept when it comes to serving as an effective means in evaluating school programs and individual performance. Previous examples of incorporating student performance data to enhance teaching and learning experienced success because of strong buy-in and support (Fishman et al., 2003; Pinkerton 2011; Reid, 2006). Strong teacher buy-in increases the likelihood that evaluations with student performance data will experience similar outcomes. One way to increase teacher buy-in and support is to identify how teachers perceive the use of student performance data in their evaluations. Instructional leaders
can then address teacher-perceived issues with the evaluation program to adjust the evaluation program, educate stakeholders, and improve the overall evaluation process.

**Statement of Purpose**

The purpose of this study is to identify how teachers perceive the use of student performance data in their teaching evaluations. Since student performance and growth is a fundamental responsibility of teachers, the use of student performance data is an appropriate performance standard in a teacher’s evaluation program. Student performance data represents a powerful mechanism in recognizing effective teachers, identifying successful instructional strategies and developing meaningful professional development (Fishman et al., 2003). The ability to achieve any of these intended outcomes relies heavily on the ability of teachers and evaluators to embrace the power of student performance data to evaluate teaching and, ultimately, improve student learning. Therefore, it is incumbent upon all parties to recognize what teachers purport as the potential liabilities and benefits associated with the utilization of these data in evaluations. This greater understanding of teacher perceptions may translate to increased levels of teacher buy-in and improve the probability for more effective teacher evaluations that improve teaching and learning.

**Research Questions**

1. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the propriety standard of the Joint Committee Personnel Evaluation Standards?

2. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the utility standard of the Joint Committee Personnel Evaluation Standards?
3. To what degree do teachers perceive the use of student performance data in their
teacher evaluation as aligned with the feasibility standard of the Joint Committee
Personnel Evaluation Standards?

4. To what degree do teachers perceive the use of student performance data in their
teacher evaluation as aligned with the accuracy standard of the Joint Committee
Personnel Evaluation Standards?

5. Is there a significant difference in the perceptions of teachers toward the
incorporation of student performance data in their evaluation among teachers (a) with
different years of experience; (b) in union and non-union states; (c) in tested and non­
tested grades and courses, and (d) in elementary, middle, and high schools?

Significance of the Study

This study is significant for a number of reasons. There is currently a lack of research
related to the perceptions of teachers regarding the use of student performance data in their
evaluations, so this study adds to the body of knowledge related to the topic. Research strongly
reports how teacher buy-in and support plays a critical role in the long-term success of
educational reform initiatives (Bryk et al., 1999; Greene & Lee, 2006; Schneider & Bryk, 2000;
Turnbull, 2002). This study will provide instructional leaders at the state-, district- and building­
levels with important information about how teachers perceived the use of student performance
data in their evaluations. The research is particularly opportune for instructional leaders given
the increasing momentum by state policy makers across the nation to now require the use of
student performance data in teacher evaluations. Even in situations where the use of student
performance data in a teacher’s evaluation is mandated by the state department of education,
teachers possess a powerful influence on how successful this initiative will be in meeting its
intended outcomes (Apaza, 2009; Clarke, 2012; Gigante & Firestone, 2008; Greer, 2006; Lasseter, 2007; Olszewski, 2009).

The findings from this study may provide school leaders and administrators with vital feedback that allows them to preemptively address teachers’ concerns regarding the use of student performance data in their evaluation program. This information specifically documents where teachers perceive potential liabilities associated with student performance data. Instructional leaders were then able to utilize information from this study to proactively educate teachers and other relevant stakeholders as to how these perceived liabilities will be responsibly and appropriately addressed.

Teachers also identified potential benefits from the use of student performance data in their evaluation. The results from this study, therefore, may provide all parties with relevant information about the opportunities associated with this change. Administrators armed with this information may be able to expand on the merits of including student performance data in a teacher’s evaluation and garner invaluable buy-in from teachers. Learning more about how teachers perceive the evaluation process is important since evaluations have not historically had the power to enhance teaching and learning. This may also benefit teachers in understanding the rationale for including student performance data in their evaluations.

Since an evaluation is viewed by teachers as a significant part of his or her continued employment or teaching assignment, it is imperative for teachers to believe the evaluation is a reliable and valid indicator of their performance. Teachers have regularly argued against the use of student performance data because it fails to recognize the inherent differences in every classroom and every school (Kelsey, 2009; Sand, 2005). Learning more about how teachers with
varying levels of experience, at the secondary or elementary level, and in tested and non-tested
courses will provide more accurate information about these perceptions.

Justification

There has been a considerable amount of material written about teachers’ perceptions of
their evaluation program. Most of the research conducted on teachers’ perceptions sought to
discover how teachers felt about their current evaluation program and whether it was a
meaningful experience for the teacher (Breedlove, 2011; Clayton, 2008; Clemetsen, 2000; Davis,
2000; Doerr, 2012; King, 2003; Marks, 2005; Phillips, 2005; Sutton, 2008). These studies
primarily explored how teachers perceived their evaluation process primarily through the use of
qualitative studies that often utilized open-ended interviews and only marginally included some
quantitative findings. Other research has focused on teachers’ perceptions of school leadership
in the evaluation process. Tuytens and Devos (2010) conducted a thorough review of the
influence of school leadership on teachers’ perceptions of teacher evaluation policy. Their study
took the same approach that the understanding of teachers’ perceptions of new educational
policy is crucial since this perception ultimately shapes the policy’s implementation. The
difference between the two studies is that the work by Tuytens and Devos did not focus on
student performance data in the teachers’ evaluation. Despite this abundance of research on
teachers’ perceptions of their evaluation, theses studies failed to specifically address how
teachers perceive the use of student achievement data in their evaluation.

There has been some research on standards-based or performance-based evaluation
programs. These studies again did not focus their research to specifically address the perceptions
of teachers to the use of student performance data (Batchelor, 2008; Doherty, 2009; Engram,
2007; Ford-Brocato, 2004; Jederberg, 2006; Killian, 2010; Pizzi, 2009; Sand, 2005). Although
they included some references to student performance data in their survey or interview instruments, these studies did not seek to examine how teachers perceived the use of student performance data in their evaluation. These studies focused more on teachers perceptions to the use of state-mandated standards-based evaluation programs.

**Operational Definition of Key Terms**

The following key terms will be used in this study:

- *Adequate Yearly Progress (AYP):* The measure by which schools, districts and states are held accountable for student performance under Title I of the No Child Left Behind Act of 2001 (NCLB), the current version of the Elementary and Secondary Act.

- *Buy-in:* Teacher buy-in or support is the degree to which teachers perceive the program or model to be a good choice for their school or an asset to their professional development and whether they were personally motivated to make the program work and understood how the program was supposed to work to improve student learning (Turnbull, 2002).

- *Non-tested Grades and Courses:* Non-tested grades and courses are characterized as grades or courses where there is no state standardized assessment to provide student achievement data (Goe & Holdheid, 2011). Race to the Top guidance on measuring student achievement in non-tested grades and courses permits alternative measures of student learning and performance so long as they are rigorous and comparable across classrooms (United States Department of Education, n.d.).
• **Performance-Based Teacher Evaluation:** A systematic method of making professional judgments about teacher performance for the purposes of improving teacher instruction and personnel decision-making.

• **Performance Indicators:** Performance indicators are used in the evaluation system to identify, in observable behaviors, performance of the major job standards (Goe, Bell, and Little, 2008).

• **Performance Standards:** Performance standards represent specific domains in a teacher evaluation that include professional knowledge, instructional planning, instructional delivery, assessment of and for student learning, learning environment, professionalism, and student academic progress (Goe, Bell, and Little, 2008).

• **Professional development:** Professional development is a continuous endeavor by a professional to increase the knowledge of his/her craft through the processes of collaboration, reflection, teaching, and learning.

• **Student Performance Assessments / Data:** Student performance assessments are data measurements of student academic progress that are based on validated quantitative measures and provide data that reflect student performance (Goe, Bell, and Little, 2008).

• **Teacher evaluation:** A meaningful evaluation focuses on instructional quality and professional standards, and through this focus and timely feedback, enables teachers and leaders to recognize, appreciate, value, and develop excellent teaching.

• **Tenure:** Employment status granted to an employee, usually after a probationary period, indicating that the position or employment cannot be removed without just cause and only for statutorily specified reasons.
• **Tested Grades and Courses:** Tested grades or courses are defined as those grades and courses covered by the state standardized assessment under the ESEA (Elementary and Secondary Education Act) to determine student progress toward academic standards (Goe & Holdheid, 2011). Tested grades and courses afford relatively large and robust data sets that can be used to measure changes in students’ academic achievement (United States Department of Education, n.d.).

**Delimitations of the Study**

This study includes the following delimitation:

- Participation will be limited to public school K-12 teachers. Private and charter school educators will be excluded from the study.

- The degree to which student performance data are included in a teacher’s evaluation varies among states. Since there is a variance in how much student performance data is included in other state’s teacher evaluation programs, the perception of teachers in one state to the use of student performance data in their evaluation may not be similar to what a teacher would report in other areas of the nation.

**Limitations of the Study**

This study also had several limitations:

- The findings will be based on a limited number of survey responses and those survey responses may not be typical of the larger teacher population. Teachers may have a variety of reasons for participating in this study, and their responses may not have revealed their true interest or disinterest in the research. A teacher’s previous experience and/or knowledge regarding the use of student performance data in evaluations in other settings may have impacted his or her responses.
• Survey responses will rely on teachers being willing to candidly share their perceptions about including student performance data in their evaluation. The teachers' responses may be influenced by a concern for voicing objections on a state-mandated initiative.

• The measurement of student performance varies among grade levels and subject areas. Teachers' survey responses may be influenced by their feelings toward the manner or assessment in which student performance data were captured as opposed to the general feelings regarding the use of student performance data in their evaluations.

Because of these issues and the limited nature of the study, generalization to populations that differ significantly from the sample may be substantially limited.
Chapter 2: Relevant Literature

This chapter contains a review of the literature relevant to exploring the question of how and to what extent teachers perceive the use of student performance data in their evaluations. This literature review will show how the research questions from the study are related to the extant knowledge, identify the gaps in the literature and help to place this study into its broader scholarly context.

To frame the context of this investigation, the literature review will begin with an analysis of the historical impact of teacher evaluation programs. This section of the review will expressly examine how teacher evaluations have impacted teacher practice and professional development throughout America’s history. It will also identify how teacher evaluation programs have been influenced by external entities such as federal, state, and local governmental bodies. It will conclude with an examination of how the increased emphasis on accountability in education has impacted teacher evaluation programs. The second area of focus will examine how student performance data is used in a variety of education accountability programs. Information in this section will chronicle the public pressure for the acquisition and effective use of student performance data and the role of the federal and state governments in making this goal a reality. The research in this section is necessary to present because it will provide salient information regarding how student performance data is collected and used. This information will also identify how teachers feel about standardized tests and other common methods for measuring student achievement and how they perceive the effectiveness of this data in district and school-wide initiatives. This material will offer an invaluable perspective when examining how teachers perceive this same source of data as a component in their evaluations. This part of the literature review will also assess how teacher practice, professional development and
accountability has been influenced by the introduction of student performance data. The third primary focus of the literature review will provide an appraisal of teachers’ perceptions to various teacher evaluation methods. This portion of the study will specifically examine how teachers perceived evaluations that included the following components: classroom observation, principal evaluation, instructional artifact, portfolio, teacher self-report measure, study survey and value-added student data. A thorough analysis of relevant teacher perceptual studies that document how teachers perceive other evaluation programs will provide a rich background base of information on how teachers regard various evaluation programs. This information will seek to identify teacher perceptions that appear consistently regardless of the evaluation program and those perceptions that are specific to individual evaluation components.

**Historical Overview of Teacher Evaluation Programs**

Teacher evaluation programs closely parallel the social, economic and political influences of the era. In order to understand contemporary teacher evaluation systems, it is critical to have a thorough understanding of the evolution of evaluation programs throughout America’s history. The focus of teacher evaluations has altered greatly over time. How to identify what constitutes an effective teacher has varied as much as the evaluation instruments used by evaluators. Understanding the evolution of teacher evaluation programs helps explain why various programs have been practiced over time and why a return to student performance data in the evaluation process is being pursued today.

**Community Accountability**

America’s education system dates back to the colonial era of the 18th century. Education was not considered a professional discipline or field of study during the 1700s. This fact contributed to the absence of an effective or standard evaluation program in the colonies’ first
public schools (Marzano, Frontier, & Livingston, 2011). Since public schools were initially designed to provide religious instruction, the supervision of school and teachers was naturally conducted by clergy members (Glanz & Sullivan, 2005; Oliva & Pawlas, 2001). In situations where the clergy members were unwilling or unable to supervise the schools, the responsibility to evaluate instruction was given to community members or government officials. Laypeople supervised the schools and young untrained teachers served their students (Ellet & Teddlie, 2003; Wirt & Kirst, 2005). The absence of consensus as to the importance or nature of pedagogical expertise caused teachers to receive a wide array of feedback that often lacked quality (Marzano et al., 2011). Although the evaluator lacked sufficient pedagogical training, the teacher’s supervisors had significant authority to determine how to evaluate instruction and the authority to hire and fire teachers (Burke & Krey, 2005). The teacher was considered a servant of the community (Marzano et al., 2011) and this era was referred to as the community accountability historical phase in education (Badilali, 1998).

Professionalization

It was not until the mid-1800s and more specifically after the Civil War when educational professionals with specific training oversaw the evaluation of teachers (Glanz, 1998; Mastillo, 2011; Minnear-Peplinski, 2009). Industrialization sparked the growth of America’s cities and subsequently the growth of more complex school systems (Marzano et al., 2011). Starting in 1890, local educational governments – primarily those in major cities - exercised strong control over schools (Wirt & Kirst, 2005). These increasingly complex school systems required more specialized instructors and the need for an instructional leader who would assume supervisory roles. This educational staffing framework soon expanded from urban centers to smaller communities (Tracy, 1995). Tracy also noted that during this era, the supervisor of instruction
required “subject area knowledge and teaching skills” more so than an understanding of the “mores of community” (p. 323). The era of allowing clergy members to solely supervise and evaluate instruction ceased to exist.

During this time period, America began to experience the introduction of a formal educational system. The mid-1800s saw the dawning of the awareness that pedagogical skills are a necessary component of effective instruction (Marzano et al., 2011). Horace Mann and Henry Barnard represented two pioneers in the establishment of education laws and curriculum development in schools. This included teacher training and the first school for teacher education. The public funding aspect of these schools increased the need for communities’ accountability.

**First Scientific Phase**

Teacher evaluation was significantly impacted by the introduction of scientific management principles. Frederick Taylor was a pioneer of scientific management theory and became an outspoken proponent of incorporating these principles in educational reform. Taylor believed that education, like other systems, needed to be grounded on the need for increased efficiency and management. Taylor initially won support from engineers and business owners. His principles were soon incorporated into higher education courses and also started resonating with public school officials, policy makers and educational leaders (Marzano et al., 2011). Following Taylor’s lead, Edward Thorndike led a similar educational reform movement to instill measurement as the ultimate tool for a more scientific approach to education (Marzano et al., 2011). Ellwood Cubberley remarked in his 1929 book, *Public School Administration*, that students were similar to raw products that needed to be “shaped and fashioned into products to meet the various demands of life” (p. 338). Schools now began to follow the factory-model of production and supervision.
In a related reform movement during this era, William Wetzel championed the notion that standardized tests should be used to measure the effectiveness of teachers and schools (Marzano et al., 2011). The debate as to whether schools were to be measured by their concrete outputs or by less measurable social development standards was in full swing in the early 20th century. Guba and Lincoln (1985) labeled this era the first real generation of technical evaluation, because standardized test results became more commonplace in the evaluation of teachers. Standardized testing began gaining momentum as an objective measure of student progress and, by association, a measurement of teacher effectiveness.

The industrial revolution during the late 19th century stressed a need for structure and organization in the workplace. This need for discipline carried over to the schoolhouse and significantly impacted the design and implementation of teacher evaluation programs (Castillo, 2005). Evaluation programs implemented during this era were, therefore, primarily utilized to inspect teachers using a strict evaluation checklist (Kelehear, 2006; Nolan & Hoover, 2008). Whether it was labeled as bureaucratic supervision (Glanz, 1998) or professionalization (Badiali, 1998), objectivity was beginning to replace subjectivity in teacher evaluations.

Human Relations

Teachers witnessed yet another change in the substance of their evaluation programs during the early decades of the 20th century. The introduction of performance standards became more frequent and these standards were largely based on models from the military, business and government models (St. Maurice & Cook, 2005). Starting in the 1930s, teacher evaluation programs shifted their focus from inspection to supervision with an overlying purpose of improving instructional practices (Badiali, 1998; Sullivan & Glanz, 2000). Another change in the evaluation process during this era was an increasing value of formative assessment of
teachers (Sullivan & Glanz, 2000). The introduction of formative and ongoing assessment aligned with an emphasis by William Melchoir to attend to the emotional needs of teachers as well as their professional needs (Nolan & Hoover, 2008). Teachers began receiving more regular feedback that then cumulated in a final end-of-the-year summative report (Holland & Adams, 2002; Nolan & Hoover, 2008; Shelly, 2002).

This emphasis on formative evaluation practices and attention to the teachers' emotional needs continued until the late 1950s. Sputnik's launch in 1957 followed by the publication of the "Coleman Report" in 1966 renewed the outcry for more meaningful and standardized teacher evaluations to ensure all students were being taught by highly effective instructors (Marzano, 2003a). The launch of the Soviet satellite prompted a public and government call for advanced rigor that necessitated the need for highly skilled educators. Americans again wanted an education system that emphasized competencies. Assurances of equity in America's public schools documented in the Coleman Report forced school districts to develop stronger methods for ensuring a quality teacher was present in every child's classroom. More standardized and performance-driven evaluation programs were seen as the only solution to ensure academic rigor and equity of teacher placement were present in all of America's classrooms.

Clinical Supervision

Teacher evaluations were heavily influenced in the 1970s by the scholarly work conducted at the Harvard School of Education. Professors Cogan and Goldhammer introduced a clinical supervision model that included pre- and post-observation conferences and an emphasis on collaboration between the teacher and the evaluator (Marzano et al., 2011; Minnear-Peplinsky, 2009). The model was described by Goldhammer as being analogous to the supervisory practices used by medical professionals in teaching hospitals (Goldhammer, 1969).
Goldhammer specifically noted that the evaluation was an ongoing process where observation and discussion drove both the teacher and the supervisor to higher levels of growth and effectiveness (1969, p. 54). The model included the following five phases: pre-observation conference, classroom observation, analysis, supervision conference, and analysis of the analysis (Blasé & Blasé, 2004).

This model endured for many decades and experienced slight variations to meet the changing societal demands of education and teachers (Blasé & Blasé, 2004; Kelehear, 2006). One variation that did develop over time was the movement away from the rich, trusting dialogue envisioned by Goldhammer to a more ritualistic set of steps to be followed by school administrators (Marzano et al., 2011). Although the clinical supervision model was not always implemented with fidelity in the manner envisioned by Goldhammer, a study by Bruce & Hoehn in 1980 found that nearly 90% of all school administrators practiced some form of the clinical supervision framework during the late 20th century. The clinical supervision model clearly became an accepted evaluation model for a number of school districts.

**Technical Model**

More contemporary changes in teacher evaluation practices came as a result of increased governmental pressure for accountability. The publication of *A Nation at Risk: The Imperative for Educational Reforms* by the National Commission on Excellence in Education in 1983 exposed a number of deficiencies in the public school sector. The report noted that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity" (National Commission on Excellence in Education, 1983, p. 5). Fears of mediocrity extended to how teachers were evaluated. The concerns made public in this document were soon reinforced by other studies. The Third International Mathematics and Science Study (TIMSS) conducted in
1996, for example, reinforced the need for heightened levels of achievement and accountability in America’s schools. The TIMSS chronicled an imperative need for America to match the achievement of school systems in other countries in an increasingly global and competitive marketplace. These reports reinforced the public’s desire for greater accountability in America’s schools. This pressure for stronger standards ultimately extended to teacher evaluations.

The work of Madeline Hunter provided a new framework for teacher evaluation that responded to these mandates. Hunter introduced seven elements of an effective lesson that included the following: anticipatory set; objective and purpose; input; modeling; checking for understanding; guided practice; and independent practice (Marzano et al., 2011). Hunter’s seven elements quickly became the prescription for teacher evaluation in many states (Fehr, 2001). This framework for identifying “mastery teaching” became the talking points for the pre-conference, observation and post-conference during a teacher’s evaluation (Marzano et al., 2011).

**Reflective Supervisory Models**

More developmental and reflective models of teacher evaluation surfaced in the mid-1980s. William Glatthorn, Thomas McGreal, and Carl Glickman all introduced alternative perspectives in teacher evaluation that emphasized professional reflection. Glatthorn proposed the theory that teachers should be afforded the opportunity to personally develop their evaluation. This evaluation style further proposed differentiated opportunities for professional growth for teachers based on their strengths and weaknesses (Glatthorn, 1984). McGreal noted that evaluations should also be differentiated based on a teacher’s tenure and the nature of the evaluation. He maintained that evaluations should range from an intensive evaluation program for high-stakes decisions such as continued employment to a standard evaluation program
designed for quality assurance (Marzano et al., 2011). Although Glickman’s contributions during this era were more focused on supervision than evaluation, his work significantly influenced both aspects. Glickman maintained the most important goal of supervision was to improve instruction (1988). Through direct assistance to teachers, group development, professional development, curriculum development, and action research, Glickman argued that supervisors could dramatically impact teaching and learning (1988). A common theme in all of these evaluation programs was a commitment to differentiated, collaborative and reflective processes. These programs also included a focus on constructive dialogue between the teacher and supervisor that translated to meaningful professional development and growth (Marzano et al., 2011).

Technical Two

While debate in the 1980s continued as to what approach to evaluation was best, a seminal study surfaced as to what types of evaluation practices were actually taking place in America’s public school systems. The RAND group’s report, Teacher Evaluation: A Study of Effective Practices, found evaluation programs as being didactic and formulaic in nature (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). The current evaluation practices included an emphasis on developmental and reflective narratives that were insufficient in providing valid measurements of teacher effectiveness (p. 16). The study found four common problems associated with teacher evaluations. These problems included a lack of uniform evaluation practices, teacher resistance to feedback, principals lacking the resolve and competence to evaluate accurately, and lack of training for evaluators (Wise et al., 1984).

Charlotte Danielson’s work in the mid-1990s responded to this report’s findings and introduced yet another framework for teacher evaluation. “Danielson sought to capture—in its
full complexity—the dynamic process of classroom teaching” (Marzano et al., 2011). The four domains in her model included: planning and preparation, the classroom environment, instruction and professional responsibilities (Danielson, 2007). Danielson reported that the intent of the framework was to honor the complexity of teaching, constitute a language for professional conversation, and provide a structure for self-assessment and reflection on professional practice (2007). “The level of specificity supplied in the Danielson model provided the foundation for the most detailed and comprehensive approach to evaluation to that time” (Marzano et al., 2011).

**Student Achievement**

The 21st century has seen a renewed emphasis on student achievement in teacher evaluations. Evaluation systems that stress student gains in learning in addition to observations of classroom instruction were the focal point of research work presented by Stronge and Tucker (2005). Their work focused on evaluation programs that incorporated student learning gains. They reviewed the supervisory systems in four different school districts that used student data on instructional practices and learning gains. This study documented how “student achievement can, and indeed should be, an important source of feedback on the effectiveness of schools, administrators, and teachers” (Tucker & Stronge, 2005, p. 102). Stronge and Tucker built on the empirical work of Bill Sanders and others which demonstrated the quantifiable impact effective teachers have on student achievement (Sanders & Rivers, 1996). Subsequent research conducted in Dallas, Texas, further reported that the harmful effects from an underachieving teacher could not be fully remediated for at least three years (Mendro, 1998). The results demonstrated that teachers clearly make a difference and, therefore, student achievement must be factored into a teacher’s evaluation. Evaluation had too often relied on observing the act of teaching as opposed
to focusing on the results of teaching (Tucker & Stronge, 2005). Evaluation needed to return to
the premise that “educator accountability for student progress in learning goes hand-in-hand with
the social contract that assigns responsibility for education to schools” (Schalock, 1998, p. 237).
A movement in teacher evaluation to reintroduce objectivity and accountability began as the
instruments, and data to fairly evaluate teachers on student achievement became more readily
available and validated.

Two subsequent reports reiterated the emphasis on student achievement data in their
work. Toch and Rothman’s 2008 report Rush to Judgment offered a critical review of
evaluations as being “superficial, capricious, and often don’t even directly address the quality of
instruction, much less measure students’ learning” (p. 1). The Widget Effect (Weisberg, Sexton,
Mulhern, & Keeling, 2009) similarly criticized current evaluation systems as being
“disrespectful to teachers” and indifferent to instructional effectiveness (p. 4). Teachers
deserved and warranted a more objective evaluation on the results of their instruction. The
availability of student performance data were now perceived as a viable mechanism to provide a
more constructive evaluation for teachers.

Table 1: Chronological Report of Teacher Evaluation in America

<table>
<thead>
<tr>
<th>Evaluation Era</th>
<th>Timeframe</th>
<th>Significant Contributors</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Accountability</td>
<td>Late 1700s</td>
<td></td>
<td>Since public education centered around religious instruction, clergy members and/or community leaders evaluated teachers without sufficient training</td>
</tr>
<tr>
<td>Professionalization</td>
<td>Early 1800s</td>
<td>Horace Mann, Henry Barnard</td>
<td>Superintendents and principal teachers equipped with an increased level of training on instructional practices and pedagogical skills traveled through neighboring communities</td>
</tr>
</tbody>
</table>
To evaluate instruction

<table>
<thead>
<tr>
<th>First Scientific Phase</th>
<th>Late 1800s and Early 1900s</th>
<th>Frederick Taylor</th>
<th>Scientific management philosophy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Edward Thorndike Ellwood Cubberley Evaluation based on measurement and analysis of data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>William Wetzel Measures of student learning to evaluate teachers</td>
</tr>
<tr>
<td>Human Relations</td>
<td>Post World War II</td>
<td>William Melchoir</td>
<td>Emphasis on evaluating the skills of the teacher while considering the emotional needs of the individual</td>
</tr>
<tr>
<td>Clinical Supervision</td>
<td>1960s &amp; 1970s</td>
<td>Morris Cogan</td>
<td>Focus on observation and discussion to drive effectiveness (five phase clinical process)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robert Goldhammer</td>
<td></td>
</tr>
<tr>
<td>Technical Model</td>
<td>1980s</td>
<td>Madeline Hunter</td>
<td>Observation and scripting method of evaluations (Seven elements of effective teaching); prescriptive model</td>
</tr>
<tr>
<td>Reflective Supervisory Models</td>
<td>Mid 1980s</td>
<td>William Glatthorn</td>
<td>Focused feedback and practice; Pedagogical development comes from teacher self-reflection and differentiated evaluation programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thomas McGreal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carl Glickman</td>
<td></td>
</tr>
<tr>
<td>Technical Two</td>
<td>1990s</td>
<td>Charlotte Danielson</td>
<td>Evaluation process designed around four domains: planning and preparation, the classroom environment, instruction and professional responsibilities; extremely detailed approach to evaluation</td>
</tr>
<tr>
<td>Student Achievement</td>
<td>2000s</td>
<td>James Stronge</td>
<td>Importance of incorporating student achievement as a criterion in evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pamela Tucker</td>
<td></td>
</tr>
</tbody>
</table>

**Teacher Perceptions to Teacher Evaluation Methods**

Teachers have routinely been absent from the opportunity to share their thoughts about the construction and implementation of their evaluation program (Tuytens & Devos, 2009; Zimmerman & Deckert-Pelton, 2003). As with any evaluation process, the impact from the final summative evaluation report poses a significant impact on an employee’s continued employment
or opportunity for pay-for-performance bonuses. It is because of the high-stakes nature of the evaluation that the perceptions of all relevant stakeholders need to be appropriately considered. The input from teachers can be a significant step in ensuring the process is designed correctly, meets its intended outcomes, and leads to meaningful professional development and growth.

In addition to teachers' perceptions regarding the procedures, transparency, and frequency of their evaluations, there has been a steady source of research on teacher perceptions to various teacher evaluation methods. A number of dissertation studies have focused on perception studies that range from teachers' feelings about the administrator conducting the evaluation to the manner in which evaluations determined professional development. These perception and attitudinal studies provide a valuable perspective into how teachers view various evaluation programs. Instructional leaders and policymakers are able to utilize this feedback to better evaluate teacher evaluation systems. As expected, the results from these studies document a wide range of beliefs.

Although teachers may voice varying opinions about the best manner in which to conduct evaluations and how to most effectively utilize the evaluation data, there are common themes that emerge. Some of these common themes include a strong desire for evaluation programs that are collaborative, meaningful, timely, and directly related to their assignment. Overall, there was a strong position from teachers that evaluations have the potential to be extremely helpful. Where opposition to evaluations was present, it was primarily a reaction to the style of evaluation implemented rather than to the general idea of evaluation itself (Acheson & Gall, 2003). Teachers who criticized their evaluation program often felt disengaged from the process and felt the evaluation was futile because of their lack of input and control over the accountability process (Reeves, 2004). This feeling of futility resulted from a sense that the
absence of meaningful feedback made the entire evaluation process feel like more of a formality than a true opportunity for professional growth. To illustrate this mindset further, a recent survey of 15,176 teachers in 12 districts found that nearly 75% of teachers had not received specific feedback on how to improve their instructional practice (Weisberg et al., 2009). This same research discovered that even more alarming was the fact that newly inducted teachers also reported they had not received feedback on any area of performance in need of improvement over the course of their first three years as teachers. Although they present specific feedback to varying evaluation methods, teachers share a resounding belief that constructive evaluations are appreciated.

**Teacher Perceptions to Procedural Elements**

School systems can either use the teacher evaluation process as a “catalyst for improving teaching and learning” or as a “meaningless bureaucratic necessity” (Davis, Ellett, & Annunziata, 2002). One of the main factors in determining if the evaluation will be an exercise in futility or an opportunity for growth rests with whether teachers are aware of all the performance standards and indicators that will influence their evaluation. In a number of studies, teachers report that they desire knowing what standards or indicators they will be evaluated against and how this evaluation will be conducted (Conley, Muncey & You, 2005; Giliya, 2006; Milanowski & Heneman, 2001; Pizzi, 2009; Sand, 2005; Sanders, 2000; Seyfarth, 2002). Without a shared understanding of the process, expectations, and goals for the evaluation program, it should not be a surprise that a teacher evaluation system that is technically flawed or with conflicting expectations for the process will guarantee failure (Stronge, 2006).

Teachers also report that they favor more transparent evaluation programs (Castillo, 2005; Feeney, 2007; Sand, 2005). Teachers specifically comment that they want to know how
they will be evaluated (Castillo 2005; Giliya, 2006; Musick, 1997; Pizzi, 2009; Sand, 2005). This need for transparency is not only preferred by teachers it is often legally required. Teacher evaluations must be legally defensible and provide to teachers both procedural and substantive due process (Joint Committee on Standards for Educational Evaluation, 2008). In the same manner valid and reliable assessments require that students are aware of what they will be assessed on, the evaluation instrument and rating scales should be available to the teacher beforehand.

The frequency of evaluations and observations is often cited by teachers as an important but inconsistent factor in their evaluation process. One of the reasons why teachers fail to have full confidence in their evaluation is that the process is often varied across the school division and within their own school (Levandowski, 2000; Marks, 2005; Sand, 2005). Teachers report that the number of times that they are informally and formally observed varies from year to year and are often dependent on, among other variables, whether they have continuing contract status or on-cycle, their number of years of experience total and in the school division, and what building administrator is assigned to them (Sand, 2005). The fact that teachers are observed or evaluated at such varying intervals creates inherent challenges in attempting to standardize or bring about further objectivity to the process. This wide range in the frequency of observations is often cited by teachers as one reason why there is some distrust in the validity of the evaluation process (Levandowski, 2000; Sand, 2005).

**Teacher Perceptions to Classroom Observations**

Classroom observations are the most frequent form of evaluation for teachers and represent a critical source of performance observation (Goe, 2008). According to Protheroe (2002), effective observation by principals can help guide a teacher’s professional development
in addition to measuring teacher competence. Observations are intended to focus on the stated performance standards approved by the school district. Formal observations are observations where the administrator is visiting for an extended period of time. These formal observations may also include a review of teacher artifacts or products and a review of pertinent student data. Depending on the nature of the teacher’s contract status (probationary versus continuing), teacher’s license (provisional versus professional), and school district policies, formal observations may occur once a year or multiple times during the school year. Informal observations are informal visits to the classroom by the building administrator. These informal observations are intended to provide more frequent and ongoing information on a wide variety of contributions made by the teacher (VDOE, 2011). Classroom walk-throughs or observations of a teacher in non-classroom settings are examples of an informal observation.

The validity of the instrument used is paramount to conducting an effective classroom observation. One example of an instrument that has been validated for its relationship to student achievement is Charlotte Danielson’s (1996) Framework for Teaching (Pianta, La Paro, & Hamre, 2006). Research concluded that a student with a teacher in the top quartile based on Danielson’s rubric would score 0.10 standard deviation higher in math and 0.125 standard deviation higher in reading than a student assigned to an instructor in the bottom quartile of the rubric (Kane, Taylor, Tyler, & Wooten, 2010).

An empirical study conducted by Ing in 2010 revealed that the teachers’ perceptions in these attitudinal studies were substantiated over a significant sample. This study described the variability of classroom observations across schools and to relate these observations to the schools’ instructional climate. The purpose of this research was to identify conditions under which classroom observations effectively improve instruction. The study used factor analysis,
latent class analysis, and regression to interpret the survey responses from 319 principals and 15,818 teachers. The study found there was no evidence that the frequency of classroom observations related to the instructional climate of the school (Ing, 2010). Although classroom observations are the most often cited means of how evaluation has been historically collected, there is no evidence from teachers that observations provide constructive feedback and opportunities to truly differentiate effective and ineffective instruction (Zimmerman & Deckert-Pelton, 2003).

Two common themes emerged from teacher perception studies on classroom observations. The first of these themes that surfaced revealed classroom observations are not perceived by teachers as serving as the most objective analysis of a teacher’s performance. Teachers routinely cited the complaint that observations were unfortunately nothing more than a stressful “dog and pony show” (Giliya, 2006; Sand, 2005; Sutton, 2008). Teachers reported that the substance of these often announced evaluations was marginalized because it failed to provide an accurate report of what transpires in the classroom on a more routine basis (Castillo, 2005; Colby et al., 2002; Giliya, 2006; Levandowski, 2000). Teachers and administrators each perform their assigned role in the evaluation process and not surprisingly very few teachers are rated below satisfactory level (Westberg et al., 2009). These perceptions are aligned to research that suggested most observations did little to improve practice or instruction (Peterson, 2000) and can become “little more than a time-consuming charade” (Stronge & Tucker, 2003, p. 6).

A second theme that emerged from teacher perception studies was a general consensus that observations failed to substantially lead to detailed plans for professional growth. Teachers have long distrusted observations and evaluations as a punitive exercise that often seeks to reinforce compliance as opposed to professional growth (Conley & Glasman, 2008; Danielson,
Teachers perceive observations to often be composed of hurried visits due to the limitations of time and resources available in the school (Levandowski, 2000; Marshall, 2005; Pizzi, 2009; Sand, 2005). Teachers did not regard evaluations, based on two or three formal observations and using a checklist to determine observed strategies, instrumental in improving teaching (Colby et al., 2002; Levandowski, 2000; Sand, 2005; Sutton, 2008). This perception is supported by a large-scale study by Kauchak, Peterson, and Driscoll (1985) when they surveyed teachers in Utah and Florida. This research reported teachers found evaluations based on principal visits to have little or no effect on actual teaching practice. Many teachers in this study, they reported that this minimal impact on their teaching was due to the fact that the evaluation visits were too brief and non-rigorous in their content. The teachers’ responses support research that a reliance on a limited number of observations continues to reinforce the long-held belief that the teacher evaluation process is flawed (Attinello et al., 2006; Stronge, 2006; Tucker et al., 2003).

**Teacher Perceptions of Principal Evaluations**

The impact of teacher evaluations is often directly linked to the quality and instructional focus of the principal. Principals are expected to be instructional experts, to support curriculum, to provide professional development, to use data-driven decision making, to be visionary and to be able to unite the faculty to advance student achievement (Tucker, 2003). Few principals are trained to effectively use evaluations to improve teacher performance, and even fewer principals have time to evaluate every teacher thoroughly or follow up with the teacher appropriately (Jacob & Lefgren, 2008; Toch & Rothman, 2008; Weisberg, Sexton, Mulher, & Keeling, 2009). It is not therefore surprising to hear that teachers report that the success of evaluations systems is
highly dependent upon the leadership in their school (Bigham and Reavis, 2001; Zimmerman, 2003).

Teacher perception studies also reported that there is a sizable population of teachers who believe their building administrator lacks the time, content expertise or desire to complete a thorough and comprehensive evaluation (Levandowski 2000; Sand 2005). Teachers also suggested that principals often lacked the preparation and information to conduct meaningful observations (Looney, 2011; Sutton, 2008; Zimmerman & Deckert-Pelton, 2003).

Another criticism of evaluators cited in research is that administrators possess a natural reluctance to deliver negative evaluations (Levandowski, 2000; Sand, 2005). Since all parties have contributed a considerable amount of time to the process, teachers want to be presented with constructive feedback during the post-observation or review conference (Castillo, 2005; Levandowski, 2000). The quality of this feedback is often what distinguishes a strong evaluation from a weak one. Many teachers desire feedback so long as it is connected to the enhancement of the school’s mission (Castillo, 2005; Levandowski, 2000; Marks, 2005; Sand, 2005).

In a number of recent perception studies, when teachers were asked about their confidence in the evaluator’s rating, the scores reflect general agreement (Castillo, 2005; Engram, 2007; Levandowski, 2000; Pizzi, 2009). When asked about their confidence in the evaluator’s expertise in assuming the role of instructional leader, the scores were less positive (Levandowski, 2000; Marks, 2005; Pizzi, 2009; Sand, 2005). Teachers often complained that the principal was not knowledgeable in current teaching practices and lacked the competence and resolve to evaluate accurately (Castillo, 2005; Cooper et al., 2005; Sand, 2005). This finding is supported by the research that many of the building administrators who conduct teacher evaluation

32
evaluations have minimal expertise in content areas which limited their credibility to teachers (Zimmerman & Deckert-Pelton, 2003).

Although teachers report a number of concerns about principal evaluation, they are prone to hold their administrator as a person in a more a positive light. Teachers are also apt to initially report that their evaluation process is positive (Castillo, 2005; Sutton, 2008). This optimistic response oftentimes changes when queried if the evaluation process provides substantial feedback and guidance in their development as teachers. Teachers commonly report that their principals and evaluators are good people but lack the training, time, or expertise in conducting effective evaluations (Castillo, 2005; Nordheim, 2006). The Massachusetts Partnership for Schools also found that nearly 50% of the teachers it surveyed did not believe that “administrators spend enough time in classrooms to carry out district’s expectations for evaluation” (MassPartners, 2000). Teachers remarked that when skilled principals evaluate them they find the process helpful (Castillo, 2005) and rate the experience as positive (Nordheim, 2006). Unfortunately, teachers routinely failed to consider their principal as a true instructional leader equipped to provide significant and meaningful feedback.

Teacher Perceptions to Portfolios & Document Logs

Portfolio and document logs are not as commonplace in evaluation programs as classroom observations. In school districts where these performance-based evaluation components are present, there exists a range of responses as to how effective they are to teachers. Portfolios are designed to offer teachers and evaluators another forum to provide documentation generated by the teacher to demonstrate his or her mastery of the stated performance standards. A teacher portfolio represents one of the most authentic performance assessments for teachers (Tucker et al., 2003). Portfolios are oftentimes viewed as an assessment technique that “gets
close to the activity of interest” and can be shaped to specific contexts thus reflecting what has been termed “ecological validity” (Peterson, 2000, p. 237). Examples of such protocols include the Instructional Quality Assessment done by the National Center for Research on Evaluation, Standards, and Student Testing (Matsumura, Slater, Junker, et al., 2006) and the Intellectual Demand Assignment Protocol (IDAP) developed by the Consortium on Chicago School Research (Newmann et al., 2001). The IDAP showed that students of teachers who scored higher on the instrument had learning gains on the Iowa Test of Basic Skills that were 20 percent higher than the national average. When designed and implemented correctly, teacher portfolios can represent an individualized, ongoing record of growth, and dynamic form of assessment.

Teachers share many of the same positive responses echoed by researchers about the use of portfolios and document logs. In a significant comprehensive construct validation study regarding the efficacy of portfolios for teacher evaluation and professional development, Tucker, Stronge, Gareis, and Beers (2003) discovered that these teaching artifacts are an appropriate counterbalance to the historically heavy emphasis on observations. The use of portfolios was favored by teachers as a means to increase self-reflection and professionalism (Sutton, 2008; Tucker et al, 2003; Zepeda, 2002). This same study found that portfolios are appealing to teachers because of their authentic nature, recognition of task complexity, active involvement of participants, encouragement of reflection and self-assessment, and facilitation of collaborative interaction. The teachers’ feedback in this study is strongly aligned with research stating that a stronger evaluation process includes opportunities for evaluation that extend beyond mere observations to include performance artifacts (Attinello et al., 2006; Dyers, 2001; Tucker et al. 2003). In much the same way teachers provide students with multiple opportunities to demonstrate mastery of the course material, teachers desire the same from evaluators to improve
the reliability and validity of the evaluation report. The comprehensive construct validation study supported the premise that the portfolio-based assessment system could distinguish teachers of accomplished practice and thus address one accountability purpose for teacher evaluation (Tucker et al., 2003).

The major concern expressed by teachers is related to how valid this method is in documenting teacher performance given the time and effort required to complete portfolios. Teachers reported that the time necessary for the completion of a portfolio or document log to be a major issue in the usefulness and feasibility of this evaluation component (Tucker et al., 2003; Zepeda, 2002). There is also no conclusive evidence that exists that the process of developing a portfolio and being evaluated by that system leads to improvement in teaching practices and student learning (Attinello et. al., 2006). Teachers also shared a desire for portfolios to represent a portion of their evaluation as opposed to serving as the exclusive component of their evaluation (Tucker et al., 2003). This feeling is reinforced by research that suggests portfolios are used inclusively as one source of information in a multiple data source system for evaluation (Peterson, 2000).

**Teacher Perceptions to Self-Evaluation**

Teachers overwhelmingly convey a desire to be more involved in the design and implementation of their evaluation program (Castillo, 2005; Engram, 2007; Feeney, 2007; Giliya, 2006; Pizzi, 2009; Zimmerman & Deckert-Pelton, 2003). Teachers believe they can work in tandem with the building administrator to develop a personalized evaluation program that will specifically address instructional issues at play in their classroom (Castillo 2005; Sand 2005). Teacher evaluation logs have proven to be valid, reliable, and cost-effective means to
Further add data points in a teacher's evaluation and provide direction for further professional development (Rowan, Jacob, & Correnti, 2009).

There exist few perception studies of how teachers feel about teacher self-evaluation models. This can be attributed to the fact that only a relatively small amount of school districts utilize a self-evaluation component in their evaluation program. In one extensive empirical study published by Kyriakides and Demetriou in 2007, new findings suggested that teachers considered self-evaluation as one of the most appropriate techniques of evaluation. Teachers more specifically commented that self-evaluation was far more valuable and productive than external observations (Kyriakides & Demetriou, 2007). The 175 randomly selected teachers in this study favored self-evaluation as a step in providing them with more professional autonomy through their involvement in the process of evaluation.

School districts using the self-evaluation approach have realized varying levels of success in improving teaching. When employees participate in their own evaluations, the quality and quantity of information increases, and ratings become more accurate and valid (Roberts, 2002). Although self-evaluation is a favored source for evaluation by teachers (Castillo, 2005; Sand, 2005), no relationship from self-report evaluation measures to better student performance has been significantly documented (Kumrow & Dahlen, 2002).

**Teacher Perceptions to Student Surveys**

Most teachers do not participate in evaluation programs that involve student survey components. Historically, input to decisions about school improvement has been limited to adults making them the only observers who are allowed to influence educational change. Increasing research documents the need for a more intergenerational perspective (Strom & Strom, 2009; Strom & Strom, 2011). Students are often able to identify aspects of education that
they strongly value and note the possibilities that are absent. Research also indicates that students provide accurate feedback (Marsh, 2007; Hattie, 2009). The Davis County School District (Utah) uses pupil surveys as one teacher-chosen data source for teacher evaluation. The surveys of 9,765 students were analyzed for patterns of response. Item analysis suggests pupils responded with reliability and validity (Peterson, Wahlquist, & Bone, 2000). In a similar study, 1,976 K-12 students in Wyoming, Wilkerson, Manatt, Rogers, & Maughan (2000) found that student ratings of teachers were significant predictors of student achievement in reading (p<.001) while self-ratings by teachers, principal ratings, and principal summative evaluations were not significant at even the .05 level in reading. These findings support the premise that there is convincing evidence that student ratings of teachers are worth considering for inclusion in teacher evaluation systems (Goe, Bell, & Little, 2008).

Teachers are mixed in their feelings toward student input in their evaluations. Some teachers report that the introduction of student surveys is a disincentive to introducing academic rigor (Emery, Kramer, & Tian, 2003). In another survey, teachers who felt student surveys would not influence their teaching voiced significant concerns about the accuracy of student surveys and the sample of students surveyed (Balch, 2012). Teachers who are more enthusiastic toward student surveys find the feedback from students to be helpful in their development (Musick, 1997). In a survey of teachers who introduced student surveys, seventy-five percent of the teachers found the student reports to be very or somewhat accurate and eighty percent of them indicated that student feedback would change their practice (Balch, 2012).

**Teacher Perceptions to Student Performance Data**

Where the existence of empirical studies on perceptions of teachers is notably scarce is with respect to the use of student achievement data in teacher evaluations. This is not entirely
surprising given the fact that only recently have states begun to recommend if not require some components of student achievement data in the evaluation matrix. Virginia, for example, recently completed a thorough review of the state’s teacher evaluation process and has issued a document outlining expectations for teacher evaluations that include components related to student achievement data (VDOE, 2011). Since the empirical research strongly documents a clear relationship between good instruction and student achievement, it stands to reason that student performance should be included in a teacher’s evaluation (Stronge & Tucker, 2005).

The perceptual studies of teachers who have participated in standards-based evaluation programs indicate that there are mixed feelings about the introduction of student performance to evaluations. Many teachers express initial concern about the introduction of student performance data in their evaluations. Emery and Ohanian (2004, p. 34) reported that teachers were fearful of what harm or consequences would come to them as a result of test results interpreted incorrectly by principals or district officials. Teachers also expressed concerns that this level of increased accountability placed so much pressure on teachers that many of them would resort to “teaching to the test” (Knight, 2008). This knee-jerk response caused teachers to streamline their instructions to teach to the test because they knew that their evaluations depended on these student scores (Knight, 2008). Teachers also question the value of student performance data because of concerns over the validity of the assessment (Nowak, 2009) and whether this data truly identifies outstanding educators (Milou & Bohlin, 2003).

Teachers who were more optimistic toward the inclusion of student performance data in their evaluation pointed to the opportunities associated with this evaluation format. Teachers believed that a focus on student achievement data may actually facilitate learning (Baker et al., 2010). Others believed that student performance data offered administrators and
teachers a neutral and objective source of information that can launch constructive conversations between both parties (Lyon, 2010). This type of constructive dialogue has the opportunity to significantly assist the teacher’s professional growth. Another study found that a majority of teachers favored student achievement data so long as it was put into proper perspective (Castillo, 2005).

Table 2: Summary of Evaluation Models

(Chart adapted from Little, Goe, & Bell, 2009, p. 26-29)

<table>
<thead>
<tr>
<th>Evaluation Model</th>
<th>Background Information on this Evaluation Model</th>
<th>Teacher Perspectives on this Evaluation Model</th>
<th>Significant Contributor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom Observation</td>
<td>• Information on a wide array of teacher contributions in the classroom can be collected as one piece of a more comprehensive data collection process; multiple visits are required in order to observe and provide feedback on all of the performance standards (Stronge &amp; Tucker, 2003) • Validity and reliability evidence dependent on instrument, sampling procedures, and rater training (Goe &amp; Croft, 2009)</td>
<td>• Prefer an increased frequency of observations • Value announced and unannounced observations • Label announced observations as “dog and pony shows” • Criticize observations as often hurried obligations by administrators • Question impact of observations on professional development and student achievement</td>
<td>• Castillo, 2005 • Engram, 2007 • Giliya, 2006 • Levandowski, 2000 • Marshall, 2005 • Sand, 2005 • Sutton, 2008</td>
</tr>
<tr>
<td>Principal Evaluation</td>
<td>• Principals have a unique perspective on school and context of instruction (Brandt et al., 2007) • Popular because it provides summative scores for accountability purposes,</td>
<td>• Express positive feelings toward administrators as a person • Share concerns about the training and expertise of</td>
<td>• Castillo, 2005 • Levandowski, 2000 • Looney, 2011 • Marks, 2005 • Nordheim, 2006</td>
</tr>
<tr>
<td>Portfolios &amp; Document Logs</td>
<td>Inform decisions about tenure or dismissal, identify teachers in need of remediation, or provide formative feedback to improve teachers' practice (Little et al., 2009)</td>
<td>Portfolios &amp; Document Logs</td>
<td>• Insightful data source for documenting the work teachers actually do and provide evidence of teaching excellence (Matsumura, Slater, Junker, et al., 2006) • Promotes teacher self-reflection and are a basis for two-way communication with an evaluator (Tucker, Stronge, &amp; Gareis, 2002)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Self-Evaluation</td>
<td>• Teachers judge the effectiveness and adequacy of their performance, effects, knowledge, and beliefs for the purpose of self-improvement (Airason &amp; Gullickson, 2006) • When teachers think about what worked, what did not work, and what type of changes they might make, the likelihood of improvements increases (Tucker et al., 2002)</td>
<td>Self-Evaluation</td>
<td>• Appreciate the opportunity to participate in the evaluation process (professional autonomy) • Discovered that self-evaluation was more personally valuable and productive than external observations</td>
</tr>
</tbody>
</table>
| Student Surveys | • Information gathered assists teachers in setting goals for • Information gathered assists teachers in setting goals for | Student Surveys | • Question the appropriateness | Student Surveys | • Question the appropriateness | Student Surveys | • Emery, Kramer, &
<table>
<thead>
<tr>
<th>Student Performance Data</th>
<th>of student input</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Documented relationship between effective teachers and student achievement (Munoz &amp; Change, 2007; Nye et al., 2004)</td>
<td>• Express concern about validity of student performance data</td>
</tr>
<tr>
<td>• Validated sources of data are required to achieve desired results (Tucker &amp; Stronge, 2006)</td>
<td>• Fear data will diminish value of other contributions to the school</td>
</tr>
<tr>
<td></td>
<td>• Fear this will lead to a &quot;teach to the test&quot; mentality</td>
</tr>
<tr>
<td></td>
<td>• Believe it will help recognize effective instructors and will facilitate learning and improved professional development</td>
</tr>
<tr>
<td></td>
<td>Tian, 2003</td>
</tr>
<tr>
<td></td>
<td>• Balch, 2012</td>
</tr>
<tr>
<td></td>
<td>• Musick, 1997</td>
</tr>
<tr>
<td></td>
<td>Becker et al., 2010</td>
</tr>
<tr>
<td></td>
<td>Castillo, 2005</td>
</tr>
<tr>
<td></td>
<td>Emery &amp; Ohanian, 2004</td>
</tr>
<tr>
<td></td>
<td>Knight, 2008</td>
</tr>
<tr>
<td></td>
<td>Lyon, 2010</td>
</tr>
<tr>
<td></td>
<td>Nowak, 2009</td>
</tr>
</tbody>
</table>

**Student Performance Data in Education Accountability Programs**

**Public Pressure for the Use of Student Performance Data**

School districts have routinely been criticized as data-rich but information-poor. This characterization refers to the fact that school districts have historically failed at effectively using data sources such as student performance data to drive decision-making. School districts are taking new steps to aggressively address this perception. Mounting public pressure on schools to
increase achievement has translated to the increasing use of achievement test scores to inform instruction (Henning, 2006). Student data had been collected but school districts were slow to harness the power of this information (Furlong-Gordon, 2009). The recent surge of data-driven instructional practices has altered this mindset and forced the hand of school districts to use this information to improve instruction (Peterson & Young, 2004; Sterbinsky, Ross, & Redfield, 2006).

Instructional leaders are now recognizing the value in student performance data and the power inherent in this data to improve student learning outcomes (Sevillano, 2002). At the district-level, student performance data is increasingly utilized by instructional leaders to align their district's curriculum and instruction to the state's approved curriculum. School districts use student performance data to develop a common scope and sequence aligned to the state's curriculum. Annual student performance data is then reviewed to make necessary revisions to these district curriculum guides. The belief is that a scope and sequence aligned to the state's assessment ensures students are appropriately exposed to the tested curriculum. When students are taught material aligned with test blueprints and benchmarks, students are better positioned to score well on these assessments (Sevillano, 2002). The effective dissemination of data allows schools to be better positioned to evaluate instructional practices and programs to help address specific areas of need (Larocque, 2007; Wade, 2004).

Student performance data is also a proven source of information for identifying potential gaps in instruction and academic achievement. Access to data and the effective analysis of data by key stakeholders to improve instruction is no longer a choice, but a must (Earl & Katz, 2006). An increasing amount of attention in school districts is being focused on closing the achievement gap. Many school districts fail to secure accreditation because of achievement gaps in AYP.
subgroup reporting categories (Radmir, 2012). School districts often struggle to close these achievement gaps in historically challenging subgroups that include African-American students, economically disadvantaged students, and students with disabilities (Radmir, 2012). The use of student achievement data can assist both teachers and school leaders in the identification of student achievement gaps both at the classroom and district level (Sevillano, 2002). Disaggregated data for subgroup performance and mandated AYP adds to how data can inform decision-making in educational policy (Earl & Fullan, 2003; Ediger, 2003; Isaacs, 2003).

Role of federal government.

*Nation at Risk.* It is often stated that no other government publication contributed more to educational accountability programs than the publication, *A Nation at Risk.* Although this document rang the alarm for accountability it failed to put in place a viable mechanism to enact change. In 1985, the Bicentennial Commission published a postscript to *A Nation at Risk* which cautioned Americans about the dangers of not immediately improving the education system. “If the nation wants to reduce its risk, it must upgrade the teaching profession and the conditions under which teachers practice (p. 77). This report set into place a strong reaction. Local, state, and national entities began to dictate the development of new supervision practices and measures to assess the value of teachers and their competency (Dagley & Veir, 2002; Elmore & Fuhrman, 2001). The alarm bell sounded and subsequent federal legislation sought to increasingly mandate accountability in America’s schools.

*No Child Left Behind.* The passage of the No Child Left Behind Act (NCLB) initiated a series of accountability requirements for school districts across the nation. With the enactment of the No Child Left Behind (NCLB) Act of 2001, public schools, school districts, and states were evaluated each year to determine if Adequate Yearly Progress (AYP) was met. This act was
signed into law on January 8, 2002. This new law embodied President Bush’s education reform plan and became “the most sweeping reform of the Elementary and Secondary Education Act (ESEA) since it was enacted in 1965” (Atkinson, 2004, p. 1). NCLB legislation marked the movement toward increasing accountability and the definition of achievement standards for all students (Kucerik, 2002; Linn, 2001). Mounting pressure from policy makers argued for a standardized measure of accountability that replaced subjective grading measures found in many school districts. The mindset was that since grades are not reliable indicators of achievement, some other strategy should be applied to detect student strengths, deficits, and progress (Abernathy, 2007).

The NCLB Act was based on four basic principles: stronger accountability for results, increased flexibility and local control, expanded options for parents, and an emphasis on teaching methods that have been proven to work (Atkinson, 2004, p. 1). This legislation authorizes several federal programs that aim to improve the academic performance of primary and secondary schools by increasing standards of accountability for states, school systems or districts, and schools. This act also provides parents with more flexibility in the selection of a school for their children to attend if their child’s school was identified as underperforming. At its foundation, NCLB was intended to be viewed as federal legislation which enacts the theories of standards-based education reform. This theory is based on the belief that high expectations and setting of goals will result in success for all students (Atkinson, 2004).

Two of the most immediate accountability measures from NCLB impacting teachers were the requirement for school districts to prioritize the hiring of highly qualified teachers and increase testing to measure Annual Yearly Progress (AYP) (Birman, Le Floch, & Klekotka, 2007; Koops & Winsor, 2005). The mandate to introduce high-stakes testing created a firestorm
of debate. Advocates of testing contended that the presence of high-stakes tests was “necessary to hold schools accountable, reward high performing schools, and identify failing school so that they may be targeted for extra help” (Kohn, 2000, p. 135). Proponents of NCLB legislation further argued that these tests and other heightened accountability measures yielded increases in academic achievement (Knight, 2008). While some strongly endorsed NCLB, other groups criticized NCLB for not going far enough to provide the teachers and administrators the tools needed for change. These individuals argued that NCLB policy carries an implicit mandate that the availability of this data should inform and enhance teaching practices, but the mechanisms for helping educators turn accountability data into actionable information are oftentimes lacking (Goertz & Duffy, 2003; Marzano, 2003b; Smith, 2005; Strieffer & Schumann, 2005).

While some may highlight the shortcomings of some NCLB provisions, it is apparent that NCLB produced unprecedented attention on accountability in schools and put significant pressure on school districts, administrators, and teachers for improving student achievement (Bernhardt, 2004; Wade, 2004). This level of accountability spread to the evaluation of teachers. Leaders demonstrated their accountability to the public by establishing effective evaluation systems. Instructional supervisors, for example, increasingly started evaluating teachers based on student achievement and learning because of NCLB (Judson, Schwartz, Allen, & Miel, 2008; Shelly, 2002). Teachers and administrators quickly began to see how NCLB laid the foundation for the use of student performance data in their own evaluations (Shelly, 2002).

**Race to the Top Initiative.** President Barack Obama issued a challenge to America’s governors, school boards, principals, teachers, parents and students. In his challenge to these stakeholders, his administration pledged significant Race to the Top grant funds to states that set and enforce rigorous and challenging standards and assessments and put outstanding teachers at
the front of the classroom (United States Department of Education, July 24, 2009). This report went on to comment that, “It’s time to make education America’s national mission.” Obama’s administration laid the groundwork for promoting innovation, reform and excellence in America’s Public Schools through an unprecedented $4.35 billion investment (USDOE, 2010).

The Race to the Top Fund was part of the American Recovery and Reinvestment Act of 2009 (ARRA). The initiative emphasized rigorous standards and high-quality assessments, retaining great teachers and leaders in the classroom, using data-driven decision making to improve instruction and encouraging the use of innovative and effective approaches to struggling schools. The reform package that sought to attract and keep great teachers and leaders in America’s classroom included revising teacher evaluation programs.

Many states reviewed the Race to the Top provisions and elected to revise teacher evaluation programs to take advantage of the grant’s financial incentives. The United States Department of Education (USDOE) reported that eleven states and the District of Columbia won awards in phases one and two of the program (2010). These states pledged to design and implement reformed evaluation programs that sought to improve teacher and principal effectiveness based on performance. These “rigorous, transparent and fair evaluations systems for teachers” must take into account data on student growth (USDOE, 2010). States began devising teacher evaluation programs that ultimately sought to raise expectations of students and to accelerate the pace of school reform (Peterson & Hess, 2008).

The premise behind utilizing student performance data in teacher evaluations is the documented need for ensuring every child in every state receives a high-quality education. Secretary of Education Arne Duncan remarked, “Fifty states doing their own thing does not make sense. I worry about pressure because of No Child Left Behind to dumb standards down.”
The goal of Race to the Top is to transform 5,000 of the lowest performing schools in the nation over the next five years. Secretary Duncan believes Race to the Top possesses the power to encourage states to enhance laws, build partnerships with all key stakeholders, and advance bold and creative school reforms” (United States Department of Education, 2009).

Role of state governments.

State departments of education have historically been responsible for developing performance standards which teachers were expected to teach and students were expected to learn. These standards now include rigorous assessments to determine whether the intended learning outcomes were mastered. In the 1970s, statewide testing programs became more abundant and the number of statewide testing programs quickly grew from three to forty (Knight, 2008). Statewide assessments found their way to every state. The requirements for academic success were defined and framed by these assessment standards. As the high-stakes assessment grew, educators became increasingly accountable for students' performance on these rigorous assessments.

With the passage of No Child Left Behind legislation and the subsequent Race to the Top Initiative, the national government placed increasing levels of oversight on states with respect to the delivery of education. All states were now mandated to administer annual achievement tests but were permitted to develop their own measures (Collins and Halverson, 2009; White, Loker, March, & Sockslager, 2009). The following states implemented the passage of exit exams as a graduation requirement: Alabama, Alaska, Arizona, California, Florida, Georgia, Idaho, Indiana, Louisiana, Maryland, Massachusetts, Minnesota, Missouri, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, South Carolina, Tennessee, Texas, Utah, Virginia, and Washington. Some states employed or planned to employ other exams, such as end-of-course
tests in high school that students must pass to receive credit for courses such as algebra I, geometry, language arts/literature, biology, economics, and history. The assessments were used to measure the degree of success achieved by students, teachers, schools, and school systems (Stiggins, 2002). States were provided some latitude on how to utilize the data from these assessments. Some states used these tests to make several determinations or high-stakes decisions: whether a student should be allowed to take certain courses or programs, whether the student is promoted to the next grade, or if the student will graduate from high school. Even though tests were intended to guide and improve instruction, school districts, administrators, and educators discovered that these tests also forced them to be more accountable for responding to deficiencies (Nicholas, Glass, & Berliner, 2002).

State departments of education were required to place an enormous amount of time and resources responding to the tests’ results. States were now forced to accept that these scores were the most visible indicators to the public for measuring the success of schools and teachers in raising the achievement level of students (Knight, 2008). In response to test scores, standards-based reform initiatives explicitly defined state standards the teachers should teach and what students should learn. Educators were then compelled to use these standards to plan and deliver classroom instruction. States also began to move toward the inclusion of student performance data in the evaluation of teachers and took an increasing role in defining evaluation practices by offering more guidance and recommendations in areas of evaluation practices (Knight, 2008). These recommendations included defining teacher quality, setting minimum standards for evaluator training, and requiring data collection (Hazi & Arredondo Rucinski, 2009). States followed the federal government’s lead in answering the call for more public accountability in schools.
School districts are now developing evaluation matrices that attempt to leverage the power of student performance data to complement other teacher performance domains to more effectively evaluate teachers. There are currently over 16 states that include student performance data in teacher evaluations. The percentage of how much student performance data is factored into the evaluation instrument ranges from 33-60% (NYSED, n.d.). These additional performance domains acknowledge the research and teacher perception studies that demonstrate a comprehensive assessment of the teacher’s scope of responsibilities needs to be considered in a teacher’s evaluation (Colby et al., 2002; Levandowski, 2000; Marks, 2005; Sand, 2005).

**Joint Committee on Standards for Educational Evaluation (JCSEE)**

In an effort to address the quality of educational evaluation, the Joint Committee on Standards for Education Evaluation (JCSEE) was created in 1975. The committee is composed of major professional associations concerned with the quality of evaluation. This organization created four categories for evaluation standards. These categories include utility, feasibility, accuracy, and propriety standards. These standards are designed as a guide for selecting and applying standards in specific evaluation settings. Since each evaluation “unfolds within a unique context,” these standards are designed as a guide for selecting and applying standards in specific evaluation settings (Yarbrough et al, 2011, p. xli). Evaluation stakeholders are encouraged to initiate any evaluation with a thorough review of these standards to determine extent to which the evaluation should incorporate each standard needed for overall evaluation quality (Yarbrough et al., 2011). Evaluation instruments and processes that adhere to these four standards are better positioned to provide all stakeholders with meaningful and appropriate information to improve teaching and learning.

**Utility Standards.**
The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs (Yarbrough et al., 2011). These standards focus on the need for evaluator credibility, relevant and meaningful information and processes in the evaluation, timely and appropriate communication and reporting of evaluation findings, and concern for the consequences and influence of the evaluation. The goal for the utility standards is to “increase the likelihood that the evaluation will have positive consequences and substantial influence, as needs and opportunities appear over the course of the evaluation” (Yarbrough et al., 2011, p. 8). In the case of teacher evaluations, the utility standards require the evaluator to possess characteristics that range from technical skills, professional rapport, and a willingness to work alongside all stakeholders during the evaluation process. This standard also addresses the need for explicit values for the evaluation. “When the values underpinning the evaluation are explicit, stakeholders are better able to review and help increase utility” (Yarbrough et al., 2011, p. 8). In schools and other open social systems, evaluation processes become meaningful when teachers and other participants understand the inherent values associated with the process. Teachers and instructional leaders, for example, can then use the evaluation to identify specific areas in need of meaningful professional development.

Table 3: Description of Utility Standards

<table>
<thead>
<tr>
<th>Utility Standards</th>
<th>Description of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1 Evaluator Credibility</td>
<td>Evaluations should be conducted by qualified people who establish and maintain creditability in the evaluation context.</td>
</tr>
<tr>
<td>U2 Attention to Stakeholders</td>
<td>Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.</td>
</tr>
<tr>
<td>U3 Negotiated Purposes</td>
<td>Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.</td>
</tr>
<tr>
<td>U4 Explicit Values</td>
<td>Evaluations should clarify and specify the individual and cultural values underpinning the purposes, processes, and judgments.</td>
</tr>
<tr>
<td>U5 Relevant Information</td>
<td>Evaluation information should serve the identified and emergent needs of stakeholders.</td>
</tr>
</tbody>
</table>
Feasibility standards.

A second standard focuses on the feasibility issues surrounding the evaluation. The feasibility standards are intended to increase evaluation effectiveness and efficiency (Yarbrough et al., 2011). Effective evaluation programs in schools, for example, are not disruptive to the learning environment. The time invested by the evaluator and the teacher during the evaluation process must be respected and used in a meaningful manner to maximize the limited time available to each stakeholder. Evaluations should use effective project management strategies and recognize the cultural and political interests and needs of individuals and groups (Yarbrough et al., 2011).

Table 4: Description of Feasibility Standards

<table>
<thead>
<tr>
<th>Feasibility Standards</th>
<th>Description of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 Project Management</td>
<td>Evaluations should use effective project management strategies.</td>
</tr>
<tr>
<td>F2 Practical Procedures</td>
<td>Evaluation procedures should be practical and responsive to the way the program operates.</td>
</tr>
<tr>
<td>F3 Contextual Viability</td>
<td>Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.</td>
</tr>
<tr>
<td>F4 Resource Use</td>
<td>Evaluations should use resources effectively and efficiently.</td>
</tr>
</tbody>
</table>

Propriety standards.

Propriety standards are the third of the joint committee’s standards and ensure that evaluations are proper, fair, legal and just (Yarbrough et al., 2011). This category addresses the need for transparency, full disclosure, formal agreements, and recognition of real or perceived
conflicts of interests that may jeopardize the effectiveness of the evaluation. In school settings, it is critical that teacher evaluations should promote sound education, fulfillment of institutional missions, and effective job performance (Yarbrough et al., 2011). Teacher evaluations that meet this standard identify both strengths and weaknesses of the teacher. This standard also requires that there are appropriate processes in place to ensure access to evaluation information is protected to only those individuals with a justifiable vested interest in the information.

Table 5: Description of Propriety Standards

<table>
<thead>
<tr>
<th>Propriety Standards</th>
<th>Description of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Responsive and Inclusive Orientation</td>
<td>Evaluations should be responsive to stakeholders and their communities.</td>
</tr>
<tr>
<td>P2 Formal Agreements</td>
<td>Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.</td>
</tr>
<tr>
<td>P3 Human Rights and Respect</td>
<td>Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.</td>
</tr>
<tr>
<td>P4 Clarity and Fairness</td>
<td>Evaluations should be understandable and fair in addressing stakeholder needs and purposes.</td>
</tr>
<tr>
<td>P5 Transparency and Disclosure</td>
<td>Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.</td>
</tr>
<tr>
<td>P6 Conflicts of Interests</td>
<td>Evaluations should openly and honestly identify and address real or perceived conflicts of interest that may compromise the evaluation.</td>
</tr>
<tr>
<td>P7 Fiscal Responsibility</td>
<td>Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.</td>
</tr>
</tbody>
</table>

Accuracy standards.

Accuracy standards are the fourth category and are intended to increase the dependability and truthfulness of evaluation findings (Yarbrough et al., 2011). The accuracy standards address the completeness and soundness of the information collected. In order to meet these standards, evaluations must include valid and reliable information, sound designs and analyses, and justified conclusions and decisions in order to be meaningful. Since teacher evaluations are
susceptible to unintended sources of bias due to the absence of a concrete output measure, it is critical to include defined expectations of the teacher and utilize defensible information to measure performance against these expectations. All sources of information used in a teacher evaluation must be analyzed systematically and accurately to fully develop justifiable conclusions regarding job performance (Yarbrough et al., 2011).

Table 6: Description of Accuracy Standards

<table>
<thead>
<tr>
<th>Accuracy Standards</th>
<th>Description of Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 Justified Conclusions and Decisions</td>
<td>Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.</td>
</tr>
<tr>
<td>A2 Valid Information</td>
<td>Evaluation information should serve the intended purposes and support valid interpretations.</td>
</tr>
<tr>
<td>A3 Reliable Information</td>
<td>Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.</td>
</tr>
<tr>
<td>A4 Explicit Program and Context Descriptions</td>
<td>Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.</td>
</tr>
<tr>
<td>A5 Information Management</td>
<td>Evaluations should employ systematic information collection, review, verification, and storage methods.</td>
</tr>
<tr>
<td>A6 Sound Designs and Analyses</td>
<td>Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.</td>
</tr>
<tr>
<td>A7 Explicit Evaluation Reasoning</td>
<td>Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.</td>
</tr>
<tr>
<td>A8 Communication and Reporting</td>
<td>Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.</td>
</tr>
</tbody>
</table>

Utilization of Student Performance Data in Schools

School districts around the nation are experiencing a transformation in how they evaluate teaching and learning. Accountability through testing for students, teachers, and administrators has surfaced as the key leverage point for policymakers seeking to promote educational reform (Hoffman et al., 2001). The standard measurement for evaluation is often student performance data derived from high-stakes standardized assessments. States use student performance data to determine a school and school district’s Annual Yearly Progress (AYP) accreditation. School
districts and schools receive accreditation only when student performance data exceeds minimum benchmarks. The failure to meet these benchmarks often places superintendents and building administrators on notice by the school board and community. Student performance data also impacts students themselves. High school students, for example, must demonstrate basic proficiency levels on state standardized tests in order to graduate. Where student performance data has not been regularly included is in the evaluation of teachers. School districts are responding by developing and implementing teacher evaluation programs that now include student performance data.

School districts have routinely been criticized as data-rich but information-poor. This characterization refers to the fact that school districts have historically failed at effectively using data sources such as student performance data to drive decision-making. School districts are taking new steps to aggressively address this perception. Instructional leaders are recognizing the value in student performance data and the power inherent in this data to improve student learning outcomes. At the district-level, student performance data is utilized by instructional leaders to align their district’s curriculum and instruction to the state’s approved curriculum. School districts use student performance data to develop a common scope and sequence aligned to the state’s curriculum. Annual student performance data is then reviewed to make necessary revisions to these district curriculum guides. The belief is that a scope and sequence aligned to the state’s assessment ensures students are appropriately exposed to the tested curriculum. When students are taught material aligned with test blueprints and benchmarks, students are better positioned to score well on these assessments (Sevillano, 2002). When students score well on state assessments, school districts meet AYP.
Student performance data is also a proven source of information for identifying potential gaps in instruction and academic achievement. An increasing amount of attention in school districts is being focused on closing the achievement gap. Many school districts fail to secure accreditation because of achievement gaps in AYP subgroup reporting categories (Radmir, 2012). School districts often struggle to close these achievement gaps in historically challenging subgroups that include African-American students, economically disadvantaged students, and students with disabilities (Radmir, 2012). The use of student achievement data can assist teachers and school leaders in the identification of student achievement gaps both at the classroom and district level (Sevillano, 2002).

Introduction of Student Performance Data to Teacher Evaluations

Education systems over the past three decades have seen a tremendous shift in the expectations for student achievement. Specifically, Americans have moved from expecting more to demanding more from the nation’s public school system. This includes more accountability of America’s teachers. One of the primary reasons for greater accountability of teachers is the growing empirical research that links teacher performance with student achievement (Darling-Hammond, 2000; Stronge, Ward, Tucker, & Hindman, 2008; Wenglinsky, 2004). One of the seminal studies that laid the groundwork for this belief was a 1997 study involving thousands of students that reported that the most important factor affecting student learning is the teacher (Wright, Horn, & Sanders). This study concluded that more can be done to improve education by improving the effectiveness of teachers than by any other single factor (Wright, Horn, & Sanders, 1997). Sanders and Rivers (1996), for example, noted that teacher effects on student academic gains can be seen as both cumulative and residual. As a result of their study, Sanders and Rivers (2002) reported that “for math tests, students taught by the least effective teachers for
three consecutive years would score 52 to 54 percentile points below similar students taught by the most effective teachers for three consecutive years" (p.4). These findings attracted the attention of researchers, policymakers, and other interested parties and began compelling school districts to become more information-rich by utilizing the availability of student performance data to improve how teachers are evaluated. This study found that even when compared against a multitude of other variables, the impact of the teacher trumps all other aspects. Tucker and Stronge (2005) further described the relationship between a high-performing teacher and student performance saying, "We now know empirically that these effective teachers have a direct influence in enhancing student achievement" (p. 2). Stronge (2002) reported research indicates high-performing teachers have residual positive effects on their students’ willingness to work to their potential and beyond. Consequently, low performing teachers may actually extinguish students’ interest in the subject.

The movement toward a change in teacher evaluations is grounded in extensive research and supported by an increasing number of studies. Evidence notes that teacher quality impacts student achievement gains (Gordon, Kane, & Staiger 2006; Rivkin, Hanushek, & Kain, 2000; Rockoff, 2004; Rowan, Correnti, & Miller, 2002; Wright, Horn, & Sanders, 1997). Studies report that distinctions in teacher quality account for at least seven percent of the total variation in student-measured achievement gains (Hanushek et al., 2005). A recent study in Texas documented a positive link between teacher effectiveness scores and student mathematics scores (Valenta, 2010). A similar study expanded on this finding demonstrating how student performance in mathematics and reading predicted teacher quality. The findings of this study documented how student performance, particularly in mathematics, can be a valid predictor of teacher quality (Washington, 2011). The connectedness of teacher quality to student
performance data in these studies demonstrates a firm correlation between the two variables. The limitations in these studies due to small samples sizes and relatively small data sets are likely to be mitigated as more states move to integrate student performance data in teacher evaluations.

There is also research that documents the strong correlation between a school’s success and the incorporation of student performance data. A recent study conducted in Missouri reported that a relationship was found to exist between the inclusion of criteria specific to student achievement in the performance-based teacher evaluation program and the school’s ranking on state assessment indicators (Lyon, 2010). Schools utilizing student achievement data were regularly appearing on the list of the state’s high-performing schools. Student achievement does not happen in a vacuum. Progressive instructional leaders are recognizing the power behind student performance data to establish high expectations for both teachers and students.

In addition to recognizing that teachers significantly impact student achievement, it is just as critical to identify and document how effective teachers work to obtain high performance thresholds. Historically, teacher evaluations have not been entirely successful in linking the traits of effective instruction to a teacher’s evaluation. This failure to align teacher evaluations with effective instruction hindered accountability measures. New research into the tenants of effective teaching allowed for the development of improved evaluation systems (Stronge, 2007). This increased base of knowledge of what constitutes effective instruction has allowed evaluation systems to hold teachers more accountable for student performance.

Use of Student Performance Data to Enhance Professional Development

The term professional development has varied over time. One definition states that professional development includes specific activities and programs designed to enhance the
professional growth of teachers (Fullan, 2001). Guskey (2000) defined professional
development as, “those processes and activities designed to enhance the professional knowledge,
skills, and attitudes of educators so that they might in turn, improve the learning of students” (p.
16). The National Staff Development Council (NSDC) lengthened the definition of professional
development to label it as a comprehensive, substantiated, and intensive approach to improving
teachers’ and principals’ effectiveness in raising student achievement (2010). The common
theme in all of these definitions is the expectation for professional development to enhance and
improve the skill sets of educators in order to deliver better teaching.

In addition to defining professional development, what constitutes effective professional
development has also been the subject of many years of scholarly research. Research indicates
that effective professional development programs are often characterized as job-embedded,
meaningful, ongoing and individualized activities that include teacher ownership (Finnegan &
140 professional development studies. This analysis identified some of the common elements in
the most successful professional programs were intensive levels of sustained support after the
training and concentrated efforts on the specific topics found to be critical for success. This
study also cited that “collaboration among teachers, mentoring or coaching opportunities,
reflection on teaching experiences, and support in the context of practice” were hallmarks of
professional development was most effective when it challenged teachers, occurred over time,
and focused on methods to improve student outcomes. Student performance data has the power
to construct meaningful and effective professional development programs.
Professional development programs filled with student performance data often provide committed educators with the necessary information to continually improve. Evidence indicates that teachers who receive substantial professional development aligned to student performance data can help students achieve more. For example, based on the findings of one meta-analysis, teachers who received substantial professional development tied to student performance data boosted their students' achievement about 21 percentile points, and this effect size is fairly consistent across content areas (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Regular staff development that is directly related to a teacher's job, driven by clear goals, and based on appropriate data and teacher input, is a powerful way to improve teacher effectiveness. Another three-year longitudinal study demonstrated a strong link between the implementation of stronger professional development and student achievement (Desimone, Porter, Garet, Yoon, & Birman, 2002). Empirical evidence also supports the use of student performance data to effectively gauge areas of need for professional development (Fishman et al., 2003). Teacher attitudinal and perception studies support the teachers' desire for the use of objective data to drive professional development activities (Levandowski, 2000; Sand, 2005). Teachers report in these studies that student performance data can play a critical role in developing sustained and job-embedded professional development programs. Student performance data also presents building administrators and district leaders the unparalleled opportunity to emphasize selecting professional development offerings that relate to the content area or population of students taught. This laser-like focus of professional development programs results in higher levels of student academic success (Camphire, 2001).

Principals are also discovering that student performance data can serve as a valuable tool in their own development as instructional leaders. A study in Pennsylvania reinforces previous
research that principals who regularly review and use student performance positively influence student achievement in schools (Soslau, 2009). This study reported that the professional practice of utilizing formative assessment data in math accounted for up to ten percent of the variability in math performance. Another study documents how student performance data is being introduced with initial success in principal evaluations. The conclusions from this study show that both the principals and the principals’ evaluators purport a high degree of satisfaction for the new evaluation instruments that includes student performance data (Reid, 2006). A final study documents how principals of feeder schools effectively use student performance data to positively impact student growth and enhance vertical articulation. In collaborative sessions where value-added student data were examined, principals were able to identify elements of both effective and ineffective school programs impacting student growth (Kelsey, 2009). Empirical evidence and related studies support the use of student performance data to evaluate the performance of principals and improve student achievement. One can reasonably assume that student performance data can produce the same favorable results for teacher development.

Stakeholders’ Perceptions to the Increased Use of Student Performance Data

Growing criticism of the public education system led policymakers and educators to turn toward testing and subsequent student performance data to measure higher skills and to gain support for raising standards. When tests were developed initially, they were designed to reflect curriculum frameworks or content standards. How students did on the test was supposed to show how well they had mastered that curriculum. When tests were used properly, they served as a valuable and valid tool to measure student achievement rather than to evaluate the quality of school resources and instruction (Knight, 2008).
The arguments against an increasingly reliance on high-stakes assessments are varied. Teachers complain that a singular focus on a one-time assessment mitigates the other dimensions of a child’s acquisition and demonstration of learning. Educators are increasingly citing the emphasis on testing has led to “teaching to the test.” An example of this teaching-to-the-test mentality was found in a study in Texas. This study found that teachers complained about the requirement to emphasize test-prep materials from September through March when the Texas Assessment of Academic Skills Test were given (Shepard, 2000). Overreliance on testing paradoxically compromised educational quality by leading teachers to teach to the test, focusing their classes on narrow test-taking strategies rather than on broader, conceptual material (Carpenter, 2001). Critics of standardized tests reported that the primary purpose of the tests was “to pinpoint gains at the low end of the spectrum. The tests did little if anything to measure how much students actually were learning or how advanced their skills were” (Walker, 2000, p. 4). “Standardized tests are group-administered, usually rely on a multiple-choice format, and offer little information to educators about the learning process or the child’s skills and ability to analyze or synthesize material” (Schwartz, 2000, p. 2).

Carpenter’s research further purported that high-stakes tests, if designed or implemented inappropriately, may draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve (2001). Another report found that some assessment systems that were used harmed huge numbers of students for reasons that few understood, and that harm arose from educational experts’ failure to balance the use of standardized tests and classroom assessments in the service of school improvement” (Stiggins, 2002, p. 9). In addition, when educators realized that their careers depended on increasing students’ test scores, these educators sometime resorted to practices that were unethical and
unfair, and they tended to neglect other activities and curriculum items that were not evaluated or assessed by high-stakes tests (Stiggins, 2002).

As high-stakes testing became more accepted and utilized, the testing instruments underwent increasing levels of scrutiny. This scrutiny has allowed standardized tests to become more valid and reliable and more accepted by educators. Not only has the assessment methodology improved during the last three decades, but the standardized tests on which the approach is based have also gone through a process of improvement and development through the influence of Item Response Theory (Van der Linden & Hambleton, 1997). The multiple-choice, closed-ended question formats found on most assessments have also improved to test beyond the knowledge and comprehension levels of Bloom’s Taxonomy. Other teachers have discovered that high-stakes assessments are able to appropriately recognize teachers for their effectiveness. Researchers found that component, productive, and accountable teachers who generally excelled treated state tests as “nothing more than another useful guide and motivator, with no significant change” in the way they presented lessons to their students (Mathews, 2006, p.1).

Summary

Teaching matters. In fact, teaching is the foundation for improved educational outcomes. A substantial body of research over the last 20 years provides us with an estimate of how much impact teachers have on student growth over time compared to other identifiable factors. This research demonstrated that individual teachers account for the largest differences between students at the end of any given year after controlling for the differences that students bring to the classroom at the beginning of the year (Wright, Horn, & Sanders, 1997; Rivkin, Hanushek, &
Kain, 2000; Rowan, Correnti, & Miller, 2002; Rockoff, 2004; Gordon, Kane, & Staiger 2006). Since teaching matters tremendously, the evaluation of teachers matters.

The eclectic variety of historical supervision practices has played a significant role in modern teacher evaluation (Minnear-Peplinski, 2009). Existing research documented an inconsistent focus in teacher evaluations throughout America’s history. Teachers have been evaluated on a number of different performance standards that have ranged from an emphasis on how well they instructed students on community norms to how rigorously they adhered to local curriculum. Oftentimes, the focus of teacher evaluations mirrored the social, political, and economic priorities of the era. Research in the history of teacher evaluation also demonstrated that one of the recurring themes in teacher evaluation practices is an emphasis on student achievement. Contemporary teacher evaluations are returning to this emphasis on student performance that places a high premium on student achievement data.

The literature also revealed that public pressure from governmental leaders and other policymakers represented the catalyst for these contemporary accountability measures in school reform. Heightened expectations from the public for more transparent and consistent measures of academic progress prompted an education reform movement that reintroduced an emphasis on standardized testing. Teacher evaluation specifically became “a pressing issue in education and educational reform” (Pearlman & Tannenbaum, 2003, p. 633).

The purpose of this study is to identify teacher perceptions to the use of student performance data in teacher evaluations. The extant literature demonstrates that there exists considerable information on how teachers perceive other evaluation models; however, the literature is relatively absent of teacher perceptions to the newly developed state evaluation models that emphasize student achievement. Research has found that when teachers examine
specific data about student performance and compare these results to constructive, detailed, and evidence-based feedback about their instruction, professional practice can improve substantially (Wenglinsky, 2002). Teachers and students were the stakeholders who were most directly affected by the demanding accountability issues, but they were often left unheard or had been silenced during this debate (Knight, 2008). Positive outcomes for teachers and students are inevitable when student performance data is appropriately implemented and effectively utilized in teacher evaluations. Understanding how teachers perceive the introduction of student performance data will certainly assist in the development of teacher buy-in which is a cornerstone for lasting and substantive change in schools (Bascia & Hargreaves, 2000; Sarason, 1995; Turnbull, 2002).
Chapter 3: Methodology

Teachers significantly influence a student's academic achievement (Sanders & Rivers, 1996; Tucker & Stronge, 2005). Until recently, however, student achievement data were noticeably absent from most teacher evaluation programs. Many states are redeveloping evaluation systems to include student performance data in response to increasing research demonstrating the link between effective teachers and student achievement (Darling-Hammond, 2000; Olson, 2008; Stronge, Ward, Tucker, & Hindman, 2008; Wenglinsky, 2004) and growing public pressure for the inclusion of student performance data in teacher evaluations (Henning, 2006; Peterson & Young, 2004; Sterbinsky, Ross, & Redfield, 2006; Weisberg, Sexton, Mulhern, & Keeling, 2009). The purpose of this study is to identify how teachers perceive the introduction of student performance data into their evaluations. To achieve this purpose, a quantitative study was conducted using a web-based survey as the data collection instrument. Participants were asked to identify their perceptions toward this new evaluation component.

Chapter three outlines the study's methodology including the research sample, instrumentation, data collection, data analysis, and ethical considerations. The sample section includes a description of the study participants. The instrumentation section discusses the survey tool and its validity and reliability evidence in addition to the interview question design. The data collection section describes the pilot study and the procedures proposed for the administration of the survey. The data analysis section outlines the proposed methods by which the survey data will be analyzed. The ethical consideration section documents what safeguards were utilized for study participants.
Research Questions

1. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the propriety standard of the Joint Committee Personnel Evaluation Standards?

2. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the utility standard of the Joint Committee Personnel Evaluation Standards?

3. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the feasibility standard of the Joint Committee Personnel Evaluation Standards?

4. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the accuracy standard of the Joint Committee Personnel Evaluation Standards?

5. Is there a significant difference in the perceptions of teachers toward the incorporation of student performance data in their evaluation among teachers (a) with different years of experience; (b) in union and non-union states; (c) in tested and non-tested grades and courses; and (d) in elementary, middle, and high schools?

Sample

The target population for this study is K-12 teachers in public schools in the United States. Stratified random sampling will be used to identify a national sample of 5,000 teachers to participate in the study. The use of systematic random sampling provides an efficient means of selecting participants from a large accessible population. The sample also will be stratified to ensure the participants equally represent elementary, middle, and high school teachers. Stratified
random sampling ensures that subgroups that are important to this study—in this case, teachers divided by elementary, middle, and high school grade levels—are represented in the sample (Gall, Gall, & Borg, 2007).

The survey sample will be acquired using the services of Market Data Retrieval (MDR). This company specializes in education marketing and possesses over three million K-12 teachers’ email addresses (http://www.schooldata.com/pdfs/MDR_Ed_catalog.pdf). MDR will assemble a list of 5,000 randomly selected public school teachers in the United States. The list will be stratified to ensure that elementary, middle, and high school teachers are evenly represented.

Instrumentation

A survey was developed for this study based on the research and work conducted by Joan Herman and Shari Golan on teachers’ perceptions of standardized testing and its impact on teachers and learners (1991) and Ansie Lessing on teachers’ perceptions of the value of professional development (2007). Herman and Golan’s survey instrument was adapted with written permission from the researchers and through the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) by The Regents of the University of California as supported under the Institute of Education Services (IES), U.S. Department of Education. Lessing’s survey was adapted with written permission from the author. The survey instrument contains 38 forced choice items. Each item includes a four point Likert-type scale where respondents were asked to identify if they strongly disagreed, disagreed, agreed, or strongly agreed with the statement. The survey concludes with participants responding to two open-ended questions and six demographic questions related to the research questions. Demographic questions pertain to the participant’s years of experience, location of current
employment, union membership status, whether they teach a tested or non-tested grade or course, and whether they teach at a high school, middle school, or elementary school.

The survey instrument was field tested by a panel of doctoral students and then submitted to a panel of experts in educational research. Twelve doctoral students at The College of William & Mary with varying levels of teaching, administrative, and other education-related experience participated in the first field test. The panel of experts was composed of four highly qualified experts in the educational research field. In both the field test and the expert panel review of the survey, participants reviewed the statements, directions, and format of the survey. Both groups were also testing the survey to ensure that the statements in the survey included content relevant to the study and research questions.

The survey was initially reviewed by the 12 doctoral students in an education leadership seminar class. All participants reported that the directions were clear and did not have any suggestions for how the directions could be improved. One typographical error in the directions was identified by a participant and this error was corrected. All of the participants who responded about the format of the survey favored the survey’s format and design. Participants were invited to offer suggestions as to whether any other aspects related to the teacher’s perception of student performance data in his or her evaluation should be included in this survey.

An additional aspect of this field test was to identify if the survey statements included content relevant to the study. Every statement was perceived by all participants to include content relevant to the study. Participants indicated whether they found each question to include content relevant to the study by answering yes or no to the question, “Do you believe the question is relevant to the study?” following each survey question. No participant identified any suggestions for additional aspects to be included in the survey. Finally, all of the doctoral
students indicated that, if invited, they were inclined to participate in the final version of this survey.

The panel of four research experts was then presented the revised survey and a report of feedback from the doctoral students for their consideration and review. For the purposes of this study, an expert is defined as an individual with extraordinary insight into the population and/or subject beyond what a member of the population under study or participant in the phenomenon being investigated might have (Ramirez, 2002). The four expert reviewers all have a doctoral degree in Educational Policy and Leadership and possess considerable experience in the design, implementation, and review of scholarly research. Dr. Min Sun, Assistant Professor in Educational Policy and Quantitative Methods at Virginia Polytechnic Institute and State University, Dr. Leslie Grant, Assistant Professor of Education at The College of William & Mary, Dr. Marco Munoz, Evaluation Specialist in the Data Management, Planning, and Program Evaluation Services Division at the Jefferson County Public Schools (Louisville, Kentucky), and Dr. Virginia Tonneson, educational consultant and recent contributor to Virginia Department of Education State-wide Teacher Evaluation Project, served as reviewers. This expert review helped to determine the credibility, conformability, and dependability of the survey. Recommendations and changes indicated by the reviewers were incorporated into the final survey. Research in the development of valid and reliable surveys documents that expert reviewers have the ability to improve in surveys by providing input on the content of the questionnaire, importance and meaningfulness of question areas to research aims, and wording and terminology of items (Dillman, 2002).

One comment that appeared in three of the four reviews of the survey was the need to include more specificity about how the numbers on the Likert-type scale corresponded to levels
of agreement or disagreement with the statement. A revision to the survey was made to ensure there was no ambiguity in how the numerical rankings corresponded to Strongly Disagree, Disagree, Agree, and Strongly Agree. The information included in Table 3.1 denotes questions in the survey that were identified by at least one of the four members as needing clarification or revision. Questions not noted in this table received 100% agreement that the question was worded clearly and relevant to the research study. The final version of the survey is presented in Appendix A.

Table 7: Panel of Experts' Feedback on Survey

<table>
<thead>
<tr>
<th>Original Question</th>
<th>The question was worded correctly?</th>
<th>The question was relevant?</th>
<th>Revised Question Wording</th>
</tr>
</thead>
<tbody>
<tr>
<td>#9 I believe the use of student performance data as one performance standard in my evaluation is a responsible use of student assessment data.</td>
<td>75%</td>
<td>100%</td>
<td>I believe the use of student performance data as one performance standard in my evaluation is an appropriate use of student performance data.</td>
</tr>
<tr>
<td>#10 I believe my evaluation should include data on how my students performed on appropriate and valid performance assessments.</td>
<td>75%</td>
<td>100%</td>
<td>I believe my evaluation should include data on how my students performed on valid performance assessments.</td>
</tr>
<tr>
<td>#12 I believe the use of student performance data will negate other performance standards and variables that impact teaching and learning in my classroom.</td>
<td>50%</td>
<td>100%</td>
<td>I believe the use of student performance data will negate other performance standards that impact teaching and learning in my classroom.</td>
</tr>
<tr>
<td>#15 I believe the use of student performance data in teacher evaluations will help administrators accurately evaluate teaching performance.</td>
<td>75%</td>
<td>100%</td>
<td>I believe the teacher evaluation process will be more meaningful to me with the use of student performance data.</td>
</tr>
<tr>
<td>#23 I believe the previous evaluation process could accurately identify for my administrator specific content areas where I can improve as a teacher.</td>
<td>75%</td>
<td>75%</td>
<td>I believe the previous evaluation process provided me with specific feedback as to where I can improve as a teacher.</td>
</tr>
<tr>
<td>Question</td>
<td>75%</td>
<td>100%</td>
<td>This question was removed from the survey prior to the expert panel review.</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-----</td>
<td>------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>#25 I believe the previous evaluation process could accurately identify whether my instruction was accurately aligned with the curriculum.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#26 I believe the previous evaluation process prevented the identification of curricular concerns at previous grade levels that may indirectly impact my students’ performance.</td>
<td>75%</td>
<td>100%</td>
<td>I believe the previous evaluation process prevented the identification of curricular concerns at previous grade levels that may impact my students’ performance this year.</td>
</tr>
<tr>
<td>#29 I believe the previous evaluation process provides school administrators with sufficient information to make informed personnel decisions.</td>
<td>75%</td>
<td>75%</td>
<td>I believe the previous evaluation process provides school leaders with sufficient information to make informed personnel decisions.</td>
</tr>
<tr>
<td>#31 I believe the use of student performance data in my evaluation will lead to the development of meaningful content or grade-level specific professional development for me as a teacher.</td>
<td>75%</td>
<td>100%</td>
<td>I believe the use of student performance data in my evaluation will lead to the development of meaningful professional development for me as a teacher.</td>
</tr>
<tr>
<td>#36 I believe the use of student performance data in my evaluation will allow for the identification of curricular concerns at previous grade levels that may indirectly impact my students’ performance.</td>
<td>75%</td>
<td>75%</td>
<td>I believe the use of student performance data in my evaluation will allow for the identification of curricular concerns at previous grade levels that may impact my students’ performance this year.</td>
</tr>
</tbody>
</table>

A final panel of experts in educational research was assembled to ensure that the survey questions were aligned to each of the four evaluation standards. The three expert reviewers all have a doctoral degree in Educational Policy and Leadership and possess considerable experience in the design, implementation, and review of scholarly research. Dr. Jennifer Hindman, Coordinator at the School Leadership Institute and the School University Research Network at The College of William & Mary, Amy Colley, Assistant Superintendent of Instruction and Support Services with Poquoson, Virginia, Public Schools, and Dr. Lisa Pennycuff, Director of Accountability and Instructional Services with York County, Virginia,
Public Schools, served as reviewers. These three individuals noted which evaluation standard best correlated with each survey question. In the four instances where the three experts did not agree, the standard that was reported most often was used. As a result of this process, seven questions (originally numbered as #8, 18, 26, 29, 34, 35, and 37) were removed from the survey because they did not align with the evaluation standards or match with another question for reliability purposes. The removal of these questions also satisfied an earlier comment from the pilot group that the survey may be too long.

Table 8: Table of Specifications for Survey

<table>
<thead>
<tr>
<th>Survey Questions</th>
<th>Key Content / Construct</th>
<th>Evaluation Standard</th>
<th>Expert Panel Agreement</th>
<th>Research Question Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &amp; 15</td>
<td>The perspective of teachers on whether the use of student performance data improves the evaluation process and rights of the teachers</td>
<td>Propriety</td>
<td>67%</td>
<td>1</td>
</tr>
<tr>
<td>2 &amp; 6</td>
<td>The perspective of teachers on whether the use of student performance data informs and improves teaching</td>
<td>Utility</td>
<td>100%</td>
<td>2</td>
</tr>
<tr>
<td>3 &amp; 4</td>
<td>The perspective of teachers on whether the use of student performance data increases attention on assessment data that improves teaching performance</td>
<td>Utility</td>
<td>100%</td>
<td>2</td>
</tr>
<tr>
<td>8 &amp; 9</td>
<td>The perspective of teachers on whether they believe they should be evaluated with the use of student performance data</td>
<td>Feasibility</td>
<td>67%</td>
<td>3</td>
</tr>
<tr>
<td>5 &amp; 7</td>
<td>The perspective of teachers on whether the use of student performance data improves the accuracy of evaluations</td>
<td>Accuracy</td>
<td>100%</td>
<td>4</td>
</tr>
<tr>
<td>10 &amp; 11</td>
<td>The perspective of teachers on whether the use of student performance data will appropriately describe the context surrounding the data collected</td>
<td>Feasibility</td>
<td>100%</td>
<td>3</td>
</tr>
<tr>
<td>12 &amp; 13</td>
<td>The perspective of teachers on whether the use of student performance data appropriately evaluates the job expectations of teachers</td>
<td>Utility</td>
<td>100%</td>
<td>2</td>
</tr>
<tr>
<td>Original Question Number</td>
<td>Questions Removed from Survey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I believe the use of student performance data will reduce evaluator subjectivity in my evaluation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCSEE Standard</td>
<td>Research Question</td>
<td>Key Content / Construct</td>
<td>Survey Questions Related to Each Standard</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| Propriety      | 1                 | • The perspective of teachers on whether the use of student performance data improves the evaluation process and rights of the teachers  
• The perspective of teachers on whether the use of student performance data improves the ratings to differentiate between levels of performance | 4 questions                                      |
| Utility        | 2                 | • The perspective of teachers on whether the use of student performance data informs and improves teaching  
• The perspective of teachers on whether the use of student performance data increases attention on assessment data that improves teaching performance  
• The perspective of teachers on whether the use of student performance data appropriately evaluates the job expectations of teachers  
• The perspective of teachers on whether the use of evaluation programs without student performance data appropriately informed professional development  
• The perspective of teachers on whether the use of evaluation programs with student performance data appropriately informed professional development | 10 questions                                      |
| Feasibility | 3 | - The perspective of teachers on whether the use of student performance data is a job-embedded function of teachers  
- The perspective of teachers on whether the use of student performance data will appropriately describe the context surrounding the data collected | 4 questions |
|---|---|---|
| Accuracy | 4 | - The perspective of teachers on whether the use of student performance data improves the accuracy of evaluations  
- The perspective of teachers on whether the use of student performance data will appropriately describe the context surrounding the data collected  
- The perspective of teachers on whether the use of student performance data improves the evaluation process because the evaluation is based on justifiable and documented performance  
- The perspective of teachers on whether the use of evaluation programs without student performance data accurately and reliably evaluated their performance as a teacher  
- The perspective of teachers on whether the use of evaluation programs without student performance data accurately and reliably evaluated other teachers’ performance  
- The perspective of teachers on whether the use of evaluation programs with student performance data accurately and reliably evaluated their performance as a teacher  
- The perspective of teachers on whether the use of evaluation programs without student performance data accurately and reliably evaluated other teachers’ performance | 14 questions |

**Procedures**

Participants will receive an email from the researcher informing them of their selection to participate in the study. The email will contain a link to an online survey in which participants identify their perceptions to the use of student performance data in their teacher evaluation. A second email will be sent within one week of the first email to remind those participants who
have not yet completed the survey. Participants who have still not completed the online survey will receive a third email one week after the second email is sent. Participants will be sent a total of three emails alerting them to the study. Research indicates that 91% of data from online surveys is collected within the first 13 days (Mitra, Jain-Shukla, Robins, Champion, and Durant, 2008); therefore, appropriate reminders will be sent early to encourage a favorable response rate from participants. The survey will remain active for one month from the date when initial contact with participants is made.

**Data Analysis**

Descriptive statistics, Cronbach’s alpha, and analysis of variance (ANOVA) will be used to analyze results from the survey. The survey questions associated with each research question will be analyzed through a Cronbach’s alpha test to determine a coefficient of internal consistency. Calculations for research question five will be subjected to an ANOVA test for each demographic variable in the survey to determine significance within groups. For example, to determine if the level of school in which teachers worked significantly impacted perceptions, ANOVA will be run for calculations provided by teachers using level of school (elementary, middle, and high school) as the independent variable.

**Table 9: Data Analysis Table**

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Data Sources</th>
<th>Data Analysis Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the propriety standard of the Joint Committee Personnel Evaluation Standards?</td>
<td>Teacher Perception Survey (Questions # 1, 15, 14, &amp; 16)</td>
<td>Descriptive statistics, Cronbach’s alpha</td>
</tr>
</tbody>
</table>
2. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the utility standard of the Joint Committee Personnel Evaluation Standards?

Teacher Perception Survey (Questions # 2, 3, 4, 6, 12, 13, 19, 20, 25, & 26)

Descriptive statistics, Cronbach’s alpha

3. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the feasibility standard of the Joint Committee Personnel Evaluation Standards?

Teacher Perception Survey (Questions # 5 & 9)

Descriptive statistics, Cronbach’s alpha

4. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the accuracy standard of the Joint Committee Personnel Evaluation Standards?

Teacher Perception Survey (Questions # 7, 8, 10, 11, 18, 19, 21, 22, 23, 24, 27, 28, 29 & 30)

Descriptive statistics, Cronbach’s alpha

5. Is there a significant difference in the perceptions of teachers toward the incorporation of student performance data in their evaluation among teachers (a) with different years of experience; (b) in union and non-union states; (c) in tested and non-tested courses, and (d) in elementary, middle, and high schools?

Teacher Perception Survey (Demographic questions)

Descriptive statistics, ANOVA

### Ethical Considerations

The researcher will also receive approval for the study from The William and Mary School of Education Internal Review Committee (EDIRC) prior to conducting any research.
Submission and approval to the EDIRC is required because doctoral dissertations constitute generalizable knowledge in the fact that the abstract will be published in Dissertation Abstracts International (Gall, Gall, & Borg, 2007). Every participant’s privacy and psychological safety will be protected throughout the study. An introductory email describing the study and the ethical safeguards included throughout the duration of the study will be sent to each participant. Participants will be free to withdraw from the study at any point.
Chapter 4: Analysis of Results

This study sought to ascertain the perceptions that K-12 public school teachers have about the use of student performance data in teacher evaluations. The researcher specifically sought to determine how teachers perceived the use of student performance data in teacher evaluations with respect to the four evaluation standards presented by the Joint Committee on Standards for Educational Evaluation (JCSEE) and if certain demographic characteristics impacted those perceptions. Data were collected using a survey created by the researcher (see Appendix A) based on the four evaluation standards created by the Joint Committee on Standards for Educational Evaluation (JCSEE, 2011).

The survey was structured into two parts. The first part had respondents use a four-point Likert scale that asked for them to indicate if they strongly agreed, agreed, disagreed, or strongly disagreed to 30 statements. The second part of the survey provided participants with an opportunity to identify up to three items that they favored and up to three items that they feared about the use of student performance data in their evaluations. Demographic information was solicited in the final four items of the survey. That information included: (a) number of years experience in education; (b) whether the teacher taught under a collective bargaining agreement; (c) whether the teacher taught a tested or non-tested grade or course; and (d) level of school in which the teacher worked (elementary, middle, or high).

Research questions one through four were addressed using Cronbach’s alpha. This measure of internal consistency is used to determine how closely related a set of items are as a group. In this study, the results demonstrated the degree to which the teacher responses to questions associated to each of the four JCSEE evaluation standards correlated with one another. These questions were previously identified by an expert panel as having been associated with
each of the four evaluation standards. Research question five was answered by computing a one-way ANOVA using the Statistical Package for Social Science (SPSS) software for each tested quality. For each ANOVA, one of the demographic factors served as the independent variable. Finally, the researcher used the open-response data to determine if teachers presented any additional perspectives about what they favored and feared regarding the use of student performance data in their evaluations.

**Research Questions**

1. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the propriety standard of the Joint Committee Personnel Evaluation Standards?

2. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the utility standard of the Joint Committee Personnel Evaluation Standards?

3. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the feasibility standard of the Joint Committee Personnel Evaluation Standards?

4. To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the accuracy standard of the Joint Committee Personnel Evaluation Standards?

5. Is there a significant difference in the perceptions of teachers toward the incorporation of student performance data in their evaluation among teachers (a) with different years of experience; (b) in union and non-union states; (c) in tested and non-tested grades and courses; and (d) in elementary, middle, and high schools?
The Study

Return Rate

The study collected data in February and March 2013. Emails were sent to a stratified equal-size random sample of elementary (grades K-5) and secondary (grades 6-12) public school K-12 teachers. The researcher created an online survey and imbedded a link to the survey in an email message to prospective participants. MDR, an educational marketing company, was retained to create an email list of K-12 public school teachers evenly stratified by elementary and secondary school levels and to deploy those emails. The list contained 5,472 teacher emails (2,733 elementary and 2,739 secondary). Table 11 documents the response rate for each subgroup.

An email was sent to participants informing them of the study and requesting their participation. Each email contained a consent agreement (see Appendix B), an introductory message (see Appendix C), and a link to the online survey. The first email was delivered on February 12, 2013. A report from MDR indicated that 311 teachers opened this email (6%), 140 continued to open the link to the survey (3%), and 134 completed the survey (2%). This first email was followed up by a reminder email on February 17, 2013 that was sent to the 171 teachers who opened the first email on February 12, 2013 but who did not click on the survey link. As a result of this second email to teachers, 139 people opened the email (81%), 111 people opened the link to the survey (65%), and 32 people actually completed the survey (19%). In total, 166 out of 5,472 K-12 public school teachers fully completed the survey for a total response rate of 3%.

The rate of participation decreased as the amount of time increased from when the participant first received the email. This finding is consistent with research on online surveys.
Within the first 24 hours of the first email, 98 surveys were completed (63% of the total responses). After twenty-four hours from the initial email, 36 additional surveys were completed for an updated total of 134 (81% of the total responses). Following the second email to participants who had opened the first email, a similar response rate occurred. Again, most responses took place within 24 hours of the message being sent to participants. The response rates significantly decreased as the time increased following each email. Table 10 indicates the response rates for each of the email deployments.

This national stratified random sample yielded 166 U.S. public school teacher-participants in the survey.

<table>
<thead>
<tr>
<th>Time following Email Deployment</th>
<th>Email sent on February 12, 2013</th>
<th>Email sent on February 17, 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 hours</td>
<td>98 surveys completed</td>
<td>15 surveys completed</td>
</tr>
<tr>
<td>4 days</td>
<td>36 surveys completed</td>
<td>6 surveys completed</td>
</tr>
<tr>
<td>6 days</td>
<td>N/A</td>
<td>5 surveys completed</td>
</tr>
<tr>
<td>10 + days</td>
<td>N/A</td>
<td>6 surveys completed</td>
</tr>
</tbody>
</table>

Table 11: Homogeneity of Responses

<table>
<thead>
<tr>
<th>Level of Participants</th>
<th>Invited to Participate</th>
<th>Number Participating</th>
<th>Percent Participating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary School Teachers (grades K-5)</td>
<td>2,733</td>
<td>91</td>
<td>3.33%</td>
</tr>
<tr>
<td>Secondary School Teachers (grades 6-12)</td>
<td>2,739</td>
<td>75</td>
<td>2.74%</td>
</tr>
</tbody>
</table>

Demographic Information

The Teacher Perception Survey: Teacher Perceptions to the Use of Student Performance Data in Teacher Evaluation included four demographic items. Those items requested information on teachers’ years of experience, whether they taught under a collective bargaining agreement, whether they taught a tested or non-tested grade or course, the level of school
(elementary, middle, or high) in which they worked, and experience with an evaluation program that utilizes student performance data. One hundred and sixty-six teachers completed the survey for a response rate of 3%. An additional 30 surveys were started but were excluded from further analysis because the participant failed to fully complete the necessary demographic information at the conclusion of the survey.

**Years of Experience.** Table 12 documents the number of completed surveys by years of experience. Thirty-three of the teachers who completed the survey had fewer than five years of teaching experience. This accounted for 20% of the total survey respondents. Teachers with 6-10 years of teaching experience accounted for 44 respondents (26% of total survey sample). Teachers with more than 10 years of experience represented the highest percentage of respondents. These teachers numbered 89 and accounted for 54% of all teachers who participated in this survey. Data from the U.S. Department of Education in 2008 noted that 17% of teachers in America had three or fewer years of teaching experience, 28% of teachers had between 4 and 9 years of experience, and 54% of teachers had more than 10 years of teaching experience. The sample in this survey was closely aligned with national statistics.

**Table 12: Participants’ Years of Experience**

<table>
<thead>
<tr>
<th>Years of Experience</th>
<th>Total Teacher Sample Completing Survey N=166</th>
<th>Percent of Total Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4 Years</td>
<td>33</td>
<td>20</td>
</tr>
<tr>
<td>5-10 Years</td>
<td>44</td>
<td>26</td>
</tr>
<tr>
<td>11 + Years</td>
<td>89</td>
<td>54</td>
</tr>
</tbody>
</table>

**Collective Bargaining Agreement.** The number of teachers who responded that they taught under a collective bargaining agreement was 89 of the total sample of 166 teachers. This represented 54% of the total sample. Seventy-seven teachers noted that they did not work under
a collective bargaining agreement that represented 46% of the total sample. A 2008 report from the U.S. Department of Education documented that 53.5% of school districts in the United States operate under some form of a collective bargaining agreement (United States Department of Education, 2008). The sample in this study very closely mirrored national statistics. Table 13 identifies the breakdown of teachers’ employment under a collective bargaining agreement.

<table>
<thead>
<tr>
<th>Work under a Collective Bargaining Agreement</th>
<th>Total Teacher Sample Completing Survey N=166</th>
<th>Percent of Total Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work under a Collective Bargaining Agreement</td>
<td>89</td>
<td>54%</td>
</tr>
<tr>
<td>Do Not Work under a Collective Bargaining Agreement</td>
<td>77</td>
<td>46%</td>
</tr>
</tbody>
</table>

**Tested or Non-tested Grade or Course.** When asked whether the teacher taught a tested grade or course, 68 teachers indicated affirmative. The 68 teachers who did teach a tested grade or course represented 41% of the total survey sample. There were 98 teachers who indicated that they did not teach a tested grade or course that accounted for 59% of the total survey sample. Tested grades or courses are defined as those grades and courses covered by the state standardized assessment under the ESEA (Elementary and Secondary Education Act) to determine student progress toward academic standards (Goe & Holdheid, 2011). Tested grades and courses afford relatively large and robust data sets that can be used to measure changes in students’ academic achievement (United States Department of Education, n.d.). Non-tested grades and courses are characterized as grades or courses where there is no state standardized assessment to provide student achievement data (Goe & Holdheid, 2011). Race to the Top guidance on measuring student achievement in non-tested grades and courses permits alternative measures of student learning and performance so long as they are rigorous and comparable across classrooms (United States Department of Education, n.d.). National research indicates
that approximately 69% of teachers do not teach a tested subject or grade (Prince et al., 2009).

The teachers in this study sample fairly closely reflected national statistics. Table 14 shows the number and percentage of teachers who taught in a tested and non-tested grade or course.

Table 14: Participants in a Tested or Non-Tested Grade or Course

<table>
<thead>
<tr>
<th></th>
<th>Total Teacher Sample Completing Survey N=166</th>
<th>Percent of Total Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teach a Tested Grade or Course</td>
<td>68</td>
<td>41%</td>
</tr>
<tr>
<td>Do Not Teach a Tested Grade or Course</td>
<td>98</td>
<td>59%</td>
</tr>
</tbody>
</table>

Level of School. Elementary school teachers accounted for 55% of the survey respondents with 91 teachers participating in the survey. Twenty-eight middle school teachers (17%) and 47 high school teachers (28%) also completed the survey. Secondary teachers (middle and high school teachers) accounted for 51% of the study sample. In the 2010 Digest of Education Statistics, the number of elementary teachers was reported as 1,758,169 and the number of secondary teachers (middle and high school) was reported as 1,234,197 (U.S. Department of Education, 2010). These national numbers (59% at the elementary level and 41% at the secondary level) are very closely aligned to the breakdown in this study sample. Table 15 illustrates the breakdown of survey participants by level of school.

Table 15: Participants’ Level of School

<table>
<thead>
<tr>
<th></th>
<th>Total Teacher Sample Completing Survey N=166</th>
<th>Percent of Total Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary</td>
<td>91</td>
<td>55</td>
</tr>
<tr>
<td>Middle</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>High</td>
<td>47</td>
<td>28</td>
</tr>
</tbody>
</table>
Experience with an Evaluation Program that Uses Student Performance Data.

Another demographic question posed on this survey was how experienced the teacher was with an evaluation program that utilizes student performance data. Ninety (54%) of the teachers reported that they had no experience with such an evaluation program. Forty-two teachers (26%) indicated that they were in their first year with this type of evaluation program. Thirty-four teachers (20%) noted that they had more than one year of being evaluated with a program that utilizes student performance data. Table 16 documents the breakdown of how many teachers in the study had experience with an evaluation program that utilizes student performance data.

Table 16: Experience with an Evaluation Program that Utilizes Student Performance Data

<table>
<thead>
<tr>
<th>Experience with Evaluation Program</th>
<th>Total Teacher Sample Completing Survey N=166</th>
<th>Percent of Total Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Experience with Evaluation Program</td>
<td>90</td>
<td>54%</td>
</tr>
<tr>
<td>First Year of Evaluation Program</td>
<td>42</td>
<td>26%</td>
</tr>
<tr>
<td>More than One Year of Evaluation Program</td>
<td>34</td>
<td>20%</td>
</tr>
</tbody>
</table>

Findings for the Research Questions

Research Question One

To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the propriety standard of the Joint Committee Personnel Evaluation Standards?

The Joint Committee on Standards for Education Evaluation (JCSEE) propriety standard ensures that evaluations are proper, fair, legal and just (Yarbrough et al., 2011). This category addresses the need for transparency, full disclosure, formal agreements, and recognition of real or perceived conflicts of interests that may jeopardize the effectiveness of the evaluation. In school settings, it is critical that teacher evaluations should promote sound education, fulfillment
of institutional missions, and effective job performance (Yarbrough et al., 2011). K-12 public school teachers responded to four questions in the survey that were identified as being associated with the propriety standard. Descriptive statistics were used to identify the mean and standard deviation for each of questions. The mean for each of the four questions was extremely close to another. The range of for the means was 0.04 (2.97-2.93). It is also interesting to note that questions placed in the survey intended to verify internal consistency (question #1 & 15 and questions #14 & 16) yielded very similar mean scores. Table 17 documents the results for all four questions.

Table 17: Propriety Standard Descriptive Statistics

<table>
<thead>
<tr>
<th>Survey #</th>
<th>Description</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percent of Teachers who Agree/Strongly Agree with Question</th>
<th>Percent of Teachers who Disagree/Strongly Disagree with Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The perspective of teachers on whether the use of student performance data improves the evaluation process and rights of the teachers</td>
<td>2.95</td>
<td>0.704</td>
<td>29</td>
<td>71</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>2.95</td>
<td>0.700</td>
<td>28</td>
<td>72</td>
</tr>
<tr>
<td>14</td>
<td>The perspective of teachers on whether the use of student performance data improves the ratings to differentiate between levels of performance</td>
<td>2.97</td>
<td>0.708</td>
<td>26</td>
<td>74</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>2.93</td>
<td>0.697</td>
<td>27</td>
<td>73</td>
</tr>
</tbody>
</table>

A further test of the internal consistency of the questions correlated to the propriety standard was measured through a Cronbach’s alpha analysis. This measure of internal consistency is used to determine how closely related a set of items are as a group. For this study, the survey results demonstrated the degree to which the teacher responses to questions associated
with the propriety standard were consistent to one another. The reliability coefficient for the propriety standard subscale was 0.787. Since an alpha coefficient of 0.7 is regarded as the baseline for acceptable reliability and an alpha coefficient of 0.8 is the baseline for good reliability (George & Mallery, 2003), this alpha coefficient demonstrates that reliability is in the acceptable range. Table 18 illustrates the results from the Cronbach’s alpha analysis.

Table 18: Cronbach’s Alpha for Propriety Standard

<table>
<thead>
<tr>
<th>Cronbach’s Alpha</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.787</td>
<td>4</td>
</tr>
</tbody>
</table>

Survey participants were largely homogenous in their responses to questions associated with the propriety standard. The grand means for the 166 participants for questions related to the propriety standard was 2.95. Since 3.0 indicated disagreement, this value indicated the grand mean was extremely close to disagreement. Participants primarily selected Disagree (corresponding to a 3) followed by the selection of Agree (corresponding to a value of 2). The low standard deviation indicated that there were very few outliers in the study who responded with Strongly Agree (a value of 1) or Strongly Disagree (a value of 4). An increase in heterogeneity in the responses occurred when comparing responses of teachers with varying levels of experience with an evaluation program that utilizes student performance data. Table 19 provides information about the grand means and standard deviation for propriety standard questions by participant’s experience with an evaluation program that utilizes student performance data.
Table 19: Mean and Standard Deviation for Propriety Standard Questions by Participant's Experience with an Evaluation Program that Utilizes Student Performance Data (SPD)

<table>
<thead>
<tr>
<th>Experience with Evaluation Program</th>
<th>Total Teacher Sample Completing Survey N = 166</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Experience with Evaluation Program</td>
<td>90</td>
<td>3.14</td>
<td>0.184</td>
</tr>
<tr>
<td>First year of Evaluation program</td>
<td>42</td>
<td>2.91</td>
<td>0.466</td>
</tr>
<tr>
<td>More than One Year of Evaluation Program</td>
<td>34</td>
<td>2.49</td>
<td>0.233</td>
</tr>
<tr>
<td>Total Participants</td>
<td>166</td>
<td>2.95</td>
<td>0.569</td>
</tr>
</tbody>
</table>

Data were further analyzed to determine whether a teacher's experience with an evaluation program that utilizes student performance data accounted for significant differences within groups for the feasibility standard. ANOVA tests were run using SPSS with the significance level set at p < .05. Table 20 documents how teachers' perceptions did significantly differ based on the teacher's experience with an evaluation program that utilizes student performance data, F (2,163) = 19.426, p = 0.001.

Table 20: Propriety Standard ANOVA by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data

<table>
<thead>
<tr>
<th>Evaluation Standard</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propriety Standard</td>
<td>Between Groups</td>
<td>10.293</td>
<td>2</td>
<td>5.146</td>
<td>19.426</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>43.183</td>
<td>163</td>
<td>.265</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53.476</td>
<td>165</td>
<td>.265</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tukey post-hoc analysis revealed that this difference was attributable to teachers with one year or more experience with an evaluation program that utilizes student performance data responding significantly more favorable to propriety standard questions than did teachers in their first year of such a program and teachers with no experience. Table 21 shows the post-hoc results.
Table 21: Tukey Post-hoc Analysis for Propriety Standard by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data

<table>
<thead>
<tr>
<th>(I) Exp</th>
<th>(J) Exp</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Year</td>
<td>First</td>
<td>-.41807*</td>
<td>.11874</td>
<td>.002</td>
<td>-.6989</td>
<td>.1372</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td>-.64346*</td>
<td>.10361</td>
<td>.001</td>
<td>-.8885</td>
<td>.3984</td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>1 Year</td>
<td>.41807*</td>
<td>.11874</td>
<td>.002</td>
<td>.1372</td>
<td>.6989</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td>-.22540</td>
<td>.09618</td>
<td>.053</td>
<td>-.4529</td>
<td>.0021</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>1 Year</td>
<td>.64346*</td>
<td>.10361</td>
<td>.001</td>
<td>.3984</td>
<td>.8885</td>
<td></td>
</tr>
<tr>
<td>First</td>
<td></td>
<td>.22540</td>
<td>.09618</td>
<td>.053</td>
<td>.0021</td>
<td>.4529</td>
<td></td>
</tr>
</tbody>
</table>

* = The mean difference is significant at the 0.05 level.

Teachers were provided the opportunity to list up to three benefits and limitations associated with including student performance data in their teacher evaluation. A representative sampling of the teacher’s feedback on potential benefits related to the propriety standard indicated that student performance data would remove evaluator bias and make teachers accountable. Specifically, teachers stated that the inclusion of student performance data may benefit evaluations by “eliminating favoritism by principals during evaluations” and by “holding teachers responsible/accountable for student progress.” The teachers’ feedback on the limitations of including student performance data associated with the propriety standard centered on the fear that student performance data would become the sole source of evaluation and that administrators may misuse the data. Some of the quotes from teachers included concerns that “data now will replace everything else I do at the school which can’t be quantified in numbers,” “test scores will trump all in the evaluation,” and “my principal can’t understand scores and I am afraid it will hurt me.” The open-ended responses were coded and grouped into similar constructs. The potential benefits and limitations cited by teachers associated with the propriety standard are listed in Table 22.
Table 22: Teacher-cited Benefits and Limitations Associated with the Propriety Standard

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Removes Evaluator Bias (2)</td>
<td>Becomes Sole Source of Evaluation (14)</td>
</tr>
<tr>
<td>Makes Teachers Accountable</td>
<td>Misuse of Data by Administrator (7)</td>
</tr>
</tbody>
</table>

Research Question Two

To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the utility standard of the Joint Committee Personnel Evaluation Standards?

The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs (Yarbrough et al., 2011). These standards focus on the need for evaluator credibility, relevant and meaningful information and processes in the evaluation, timely and appropriate communication and reporting of evaluation findings, and concern for the consequences and influence of the evaluation. The goal for the utility standards is to “increase the likelihood that the evaluation will have positive consequences and substantial influence, as needs and opportunities appear over the course of the evaluation” (Yarbrough et al., 2011, p. 8).

K-12 public school teachers responded to 10 questions in the survey that were identified as being associated with the utility standard by a panel of experts. Descriptive statistics were used to identify the mean and standard deviation for each of questions. The mean for each of the 10 questions was extremely close to another. It is important to report that the responses to questions #19 & #20 were reverse-coded since they asked teachers whether an evaluation system without student performance data were of benefit (or negatively worded). The range of means for the 10 questions was 0.14 (2.52-2.38). It is also interesting to note that questions placed in
the survey intended to verify internal consistency (questions #2 & 6; #3 & 4; #12 & 13; 19 & 20; and 25 & 26) produced very similar mean scores. Table 23 documents the results for the 10 questions.

**Table 23: Utility Standard Descriptive Statistics**

<table>
<thead>
<tr>
<th>Survey Question #</th>
<th>Description</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percent of Teachers who Agree/Strongly Agree with Question</th>
<th>Percent of Teachers who Disagree/Strongly Disagree with Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>The perspective of teachers on whether the use of student performance data informs and improves teaching</td>
<td>2.47</td>
<td>0.785</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>2.52</td>
<td>0.853</td>
<td>53</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>The perspective of teachers on whether the use of student performance data increases attention on the use of assessment data that improves teaching performance</td>
<td>2.41</td>
<td>0.799</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2.38</td>
<td>0.771</td>
<td>58</td>
<td>43</td>
</tr>
<tr>
<td>12</td>
<td>The perspective of teachers on whether the use of student performance data appropriately evaluates the job expectations of teachers</td>
<td>2.39</td>
<td>0.833</td>
<td>61</td>
<td>38</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>2.39</td>
<td>0.947</td>
<td>64</td>
<td>36</td>
</tr>
</tbody>
</table>
A further test of the internal consistency of the questions correlated to the utility standard was measured through a Cronbach's alpha analysis. This measure of internal consistency is used to determine how closely related a set of items are as a group. For this study, the survey results demonstrated the degree to which the teacher responses to questions associated with the utility standard were consistent to one another. The reliability coefficient for the propriety standard subscale was 0.751 indicating adequate reliability. Table 24 illustrates the results from the Cronbach’s alpha analysis.

Table 24: Cronbach’s Alpha for Utility Standard

<table>
<thead>
<tr>
<th>Cronbach’s Alpha</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.751</td>
<td>10</td>
</tr>
</tbody>
</table>

Survey participants were largely homogenous in their responses to questions associated with the utility standard. The grand mean for the 166 participants for questions related to the utility standard was 2.43. This value indicates that participants were leaning slightly toward
agreement with most respondents selecting Agree (corresponding to a 2) followed closely by the selection of Disagree (corresponding to a value of 3). The low standard deviation indicated that there were very few outliers in the study who responded with Strongly Agree (a value of 1) or Strongly Disagree (a value of 4). An increase in heterogeneity in the responses occurred when comparing responses of teachers with varying levels of experience with an evaluation program that utilizes student performance data. Table 25 provides information about the grand means and standard deviation for utility standard questions by participant’s experience with an evaluation program that utilizes student performance data.

Table 25: Mean and Standard Deviation for Utility Standard Questions by Participant’s Experience with an Evaluation Program that Utilizes Student Performance Data (SPD)

<table>
<thead>
<tr>
<th>Experience with Evaluation Program</th>
<th>Total Teacher Sample Completing Survey N = 166</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Experience with Evaluation Program</td>
<td>90</td>
<td>2.61</td>
<td>0.294</td>
</tr>
<tr>
<td>First year of Evaluation program</td>
<td>42</td>
<td>2.28</td>
<td>0.209</td>
</tr>
<tr>
<td>More than One Year of Evaluation Program</td>
<td>34</td>
<td>2.09</td>
<td>0.289</td>
</tr>
<tr>
<td>Total Participants</td>
<td>166</td>
<td>2.43</td>
<td>0.785</td>
</tr>
</tbody>
</table>

Data were further analyzed to determine whether a teacher’s experience with an evaluation program that utilizes student performance data accounted for significant differences within groups for the utility standard. ANOVA tests were run using SPSS with the significance level set at p < .05. Table 26 documents how teachers’ perceptions did not significantly differ based on the teacher’s experience with an evaluation program that utilizes student performance data.
Table 26: Utility Standard ANOVA by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data

<table>
<thead>
<tr>
<th>Evaluation Standard</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feasibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>7.680</td>
<td>2</td>
<td>3.840</td>
<td>14.156</td>
<td>.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>44.213</td>
<td>163</td>
<td>.271</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51.893</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tukey post-hoc analysis revealed that this difference was attributable to teachers in their first year and teachers with one or more years’ experience with an evaluation program that utilizes student performance data responding more favorable to utility standard questions than did teachers with no experience. Table 27 shows the post-hoc results.

Table 27: Tukey Post-hoc Analysis for Utility Standard by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data

<table>
<thead>
<tr>
<th>(I) Exp</th>
<th>(J) Exp</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Year</td>
<td>First</td>
<td>-.19034</td>
<td>.12015</td>
<td>.255</td>
<td>-.4745 - .0939</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>-.51732*</td>
<td>.10484</td>
<td>.001</td>
<td>-.7653 - -.2693</td>
</tr>
<tr>
<td>First</td>
<td>1 Year</td>
<td>.19034</td>
<td>.12015</td>
<td>.255</td>
<td>-.0939 - .4745</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>-.32698*</td>
<td>.09732</td>
<td>.003</td>
<td>-.5572 - -.0968</td>
</tr>
<tr>
<td>None</td>
<td>1 Year</td>
<td>.51732*</td>
<td>.10484</td>
<td>.001</td>
<td>.2693 - .7653</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>.32698*</td>
<td>.09732</td>
<td>.003</td>
<td>.0968 - .5572</td>
</tr>
</tbody>
</table>

* = The mean difference is significant as the 0.05 level.

Teachers were provided the opportunity to list up to three benefits and limitations associated with including student performance data in their evaluation. A sampling of the teacher's feedback on potential benefits related to the utility standard indicated that student performance data would “make professional development more individualized—FINALLY,” “help me make better lesson plans,” “prompt us to work together toward a common goal,” and “make us really data-driven (instead of just saying we are).” The teachers’ feedback on the limitations of including student performance data associated with the utility standard noted that
this would lead to "teaching to the test," "focusing on tests instead of important lessons," and the "elimination of anything taught that doesn't appear on the state test." The open-ended responses were coded and grouped into similar constructs. The potential benefits and limitations cited by teachers associated with the utility standard are listed in Table 28.

Table 28: Teacher-cited Benefits and Limitations Associated with the Utility Standard

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guides Lesson Planning (6)</td>
<td>Teaching to the Test (16)</td>
</tr>
<tr>
<td>Identifies Student Gaps (7)</td>
<td></td>
</tr>
<tr>
<td>Enhances Personal Growth and Reflection (3)</td>
<td></td>
</tr>
<tr>
<td>Informs Professional Development (3)</td>
<td></td>
</tr>
<tr>
<td>Increases Collaboration (2)</td>
<td></td>
</tr>
<tr>
<td>Alignment to Common Core</td>
<td></td>
</tr>
<tr>
<td>Data-Driven Decision Making (2)</td>
<td></td>
</tr>
<tr>
<td>Teaches Test-Taking Skills</td>
<td></td>
</tr>
</tbody>
</table>

Research Question Three

To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the feasibility standard of the Joint Committee Personnel Evaluation Standards?

The feasibility standards are intended to increase evaluation effectiveness and efficiency (Yarbrough et al., 2011). Effective evaluation programs in schools, for example, are not disruptive to the learning environment. Evaluations should use effective project management strategies and recognize the cultural and political interests and needs of individuals and groups (Yarbrough et al., 2011). K-12 public school teachers responded to four questions in the survey that were identified as being associated with the feasibility standard by a panel of experts. Descriptive statistics were used to identify the mean and standard deviation for each of questions. The mean for each of the four questions was extremely close to another. The range of means was 0.03 (2.47-2.43). It is also interesting to note that questions placed in the survey
intended to verify internal consistency (questions #8 & 9 and #10 & 11) produced very similar mean scores. Table 29 documents the results for the four questions.

**Table 29: Feasibility Standard Descriptive Statistics**

<table>
<thead>
<tr>
<th>Survey Question #</th>
<th>Description</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percent of Teachers who Agree/Strongly Agree with Question</th>
<th>Percent of Teachers who Disagree/Strongly Disagree with Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>The perspective of teachers on whether they believe they should be evaluated with the use of student performance data</td>
<td>2.47</td>
<td>0.893</td>
<td>61</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>2.43</td>
<td>0.842</td>
<td>61</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>The perspective of teachers on whether the use of student performance data will appropriately describe the context surrounding the data collected</td>
<td>2.46</td>
<td>0.783</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>2.45</td>
<td>0.624</td>
<td>66</td>
<td>34</td>
</tr>
</tbody>
</table>

A further test of the internal consistency of the questions correlated to the feasibility standard was measured through a Cronbach's alpha analysis. This measure of internal consistency is used to determine how closely related a set of items are as a group. For this study, the survey results demonstrated the degree to which the teacher responses to questions associated with the feasibility standard were consistent to one another. The reliability coefficient for the propriety standard subscale was 0.774 indicating adequate reliability. Table 30 illustrates the results from the Cronbach's alpha analysis.

**Table 30: Cronbach's Alpha for Feasibility Standard**

<table>
<thead>
<tr>
<th>Cronbach’s Alpha</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.774</td>
<td>4</td>
</tr>
</tbody>
</table>
Survey participants were largely homogenous in their responses to questions associated with the feasibility standard. The grand mean for the 166 participants for questions related to the feasibility standard was 2.45. This value indicates that participants were leaning slightly toward agreement with most respondents selecting Agree (corresponding to a 2) followed closely by the selection of Disagree (corresponding to a value of 3). The low standard deviation indicated that there were very few outliers in the study who responded with Strongly Agree (a value of 1) or Strongly Disagree (a value of 4). This level of homogeneity extended when comparing teachers with varying levels of experience using an evaluation program that utilizes student performance data. Table 31 provides information about the grand means and standard deviation for feasibility standard questions and the mean and standard deviation for participants broken down by varying levels of experience with an evaluation program that utilizes student performance data.

Table 31: Mean and Standard Deviation for Feasibility Standard Questions by Participant’s Experience with an Evaluation Program that Utilizes Student Performance Data (SPD)

<table>
<thead>
<tr>
<th>Experience with Evaluation Program</th>
<th>Total Teacher Sample Completing Survey N = 166</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Experience with Evaluation Program</td>
<td>90</td>
<td>2.47</td>
<td>0.341</td>
</tr>
<tr>
<td>First year of Evaluation program</td>
<td>42</td>
<td>2.43</td>
<td>0.302</td>
</tr>
<tr>
<td>More than One Year of Evaluation Program</td>
<td>34</td>
<td>2.42</td>
<td>0.342</td>
</tr>
<tr>
<td>Total Participants</td>
<td>166</td>
<td>2.45</td>
<td>0.785</td>
</tr>
</tbody>
</table>

Data were further analyzed to determine whether a teacher’s experience with an evaluation program that utilizes student performance data accounted for significant differences within groups for the feasibility standard. ANOVA tests were run using SPSS with the significance level set at p < .05. Table 32 documents how teachers’ perceptions did not significantly differ based on the teacher’s experience with an evaluation program that utilizes
student performance data. The reason for similar responses may have resulted from the fact that only four questions on the survey were associated with the feasibility standard.

_Table 32: Feasibility Standard ANOVA by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data_

<table>
<thead>
<tr>
<th>Evaluation Standard</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feasibility Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.075</td>
<td>2</td>
<td>.038</td>
<td>.155</td>
<td>.856</td>
</tr>
<tr>
<td>Within Groups</td>
<td>39.651</td>
<td>163</td>
<td>.243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>39.726</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Teachers were provided the opportunity to list up to three benefits and limitations associated with including student performance data in their evaluation. The feedback from teachers related to the feasibility standard was limited. Teachers indicated that student performance data would not be able to recognize the uniqueness of school populations and therefore would not be responsive to the way their particular school or program operates. Specifically, teachers expressed concerns about how student performance data would be evaluated since “our school’s student body is different,” “my school works with special populations—it is a different school than most,” “our kids cannot score as well as kids in other schools,” and “our school is different—unique.” The open-ended responses were coded and grouped into similar constructs. There were no comments shared by teachers about potential benefits. The limitations cited by teachers associated with the feasibility standard are listed in Table 33.

_Table 33: Teacher-cited Benefits and Limitations Associated with the Feasibility Standard_

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>Does Not Recognize Uniqueness of School Population (6)</td>
</tr>
</tbody>
</table>
Research Question Four

To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the accuracy standard of the Joint Committee Personnel Evaluation Standards?

The Accuracy standards address the completeness, dependability, truthfulness, and soundness of the information collected (Yarbrough et al., 2011). In order to meet these standards, evaluations must include valid and reliable information, sound designs and analyses, and justified conclusions and decisions in order to be meaningful. Since teacher evaluations are susceptible to unintended sources of bias due to the absence of a concrete output measure, it is critical to include defined expectations of the teacher and utilize defensible information to measure performance against these expectations. All sources of information used in a teacher evaluation must be analyzed systematically and accurately to fully develop justifiable conclusions regarding job performance (Yarbrough et al., 2011).

K-12 public school teachers responded to 12 questions in the survey that were identified as being associated with the feasibility standard by a panel of experts. Descriptive statistics were used to identify the mean and standard deviation for each of questions. The mean for each of the 12 questions was extremely close to another. The range of means for the 12 questions was 0.10 (2.90-2.80). It is important to report that the responses to questions #21 & #22 and questions #23 & #24 were reverse-coded since they asked teachers whether an evaluation system without student performance data were of benefit (or negatively worded). It is also interesting to note that questions placed in the survey intended to verify internal consistency (questions #5 & 7; #17 & 18; #21 & 22; #23 & 24; #27 & 28; and #29 & 30) produced very similar mean scores. Table 34 documents the results for the 12 questions.
<table>
<thead>
<tr>
<th>Survey #</th>
<th>Description</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Percent of Teachers who Agree/Strongly Agree with Question</th>
<th>Percent of Teachers who Disagree/Strongly Disagree with Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>The perspective of teachers on whether the use of student performance data improves the accuracy of evaluations</td>
<td>2.87</td>
<td>0.808</td>
<td>33</td>
<td>68</td>
</tr>
<tr>
<td>7</td>
<td>The perspective of teachers on whether the use of student performance data improves the evaluation process because the evaluation is based on justifiable and documented performance</td>
<td>2.80</td>
<td>0.813</td>
<td>35</td>
<td>65</td>
</tr>
<tr>
<td>17</td>
<td>The perspective of teachers on whether the use of student performance data improves the accuracy of evaluations</td>
<td>2.87</td>
<td>0.813</td>
<td>32</td>
<td>68</td>
</tr>
<tr>
<td>18</td>
<td>The perspective of teachers on whether the use of student performance data improves the evaluation process because the evaluation is based on justifiable and documented performance</td>
<td>2.87</td>
<td>0.805</td>
<td>32</td>
<td>67</td>
</tr>
<tr>
<td>21</td>
<td>The perspective of teachers on whether the use of evaluation systems without student performance data accurately and reliably evaluated their performance as a teacher</td>
<td>2.87</td>
<td>0.726</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>22</td>
<td>The perspective of teachers on whether the use of evaluation programs without student performance data accurately and reliably evaluate other teachers’ performance</td>
<td>2.88</td>
<td>0.747</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>23</td>
<td>The perspective of teachers on whether the use of evaluation programs without student performance data accurately and reliably evaluate other teachers’ performance</td>
<td>2.88</td>
<td>0.768</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>24</td>
<td>The perspective of teachers on whether the use of evaluation programs without student performance data accurately and reliably evaluate other teachers’ performance</td>
<td>2.90</td>
<td>0.783</td>
<td>70</td>
<td>30</td>
</tr>
</tbody>
</table>
A further test of the internal consistency of the questions correlated to the accuracy standard was measured through a Cronbach’s alpha analysis. This measure of internal consistency is used to determine how closely related a set of items are as a group. For this study, the survey results demonstrated the degree to which the teacher responses to questions associated with the accuracy standard were consistent to one another. The reliability coefficient for the propriety standard subscale was 0.868 demonstrating good reliability. Table 35 illustrates the results from the Cronbach’s alpha analysis.

Table 35: Cronbach’s Alpha for Accuracy Standard

<table>
<thead>
<tr>
<th>Cronbach’s Alpha</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.868</td>
<td>12</td>
</tr>
</tbody>
</table>

Survey participants were largely homogenous in their responses to questions associated with the accuracy standard. The grand mean for the 166 participants for questions related to the accuracy standard was 2.86. This value indicates that participants were leaning slightly toward disagreement with most respondents selecting Disagree (corresponding to a 3) followed by the
selection of Agree (corresponding to a value of 2). The low standard deviation indicated that there were very few outliers in the study who responded with Strongly Agree (a value of 1) or Strongly Disagree (a value of 4). Where there was more heterogeneity in responses was with respect to how much experience the teachers had with an evaluation program that utilizes student performance data. Table 36 provides information about the grand means and standard deviation for accuracy standard questions by participant’s experience with an evaluation program that utilizes student performance data.

Table 36: Mean and Standard Deviation for Accuracy Standard Questions by Participant’s Experience with an Evaluation Program that Utilizes Student Performance Data (SPD)

<table>
<thead>
<tr>
<th>Experience with Evaluation Program</th>
<th>Total Teacher Sample Completing Survey N = 166</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Experience with Evaluation Program</td>
<td>90</td>
<td>3.06</td>
<td>0.259</td>
</tr>
<tr>
<td>First year of Evaluation program</td>
<td>42</td>
<td>2.80</td>
<td>0.233</td>
</tr>
<tr>
<td>More than One Year of Evaluation Program</td>
<td>34</td>
<td>2.40</td>
<td>0.306</td>
</tr>
<tr>
<td>Total Participants</td>
<td>166</td>
<td>2.86</td>
<td>0.775</td>
</tr>
</tbody>
</table>

Data were further analyzed to determine whether a teacher’s experience with an evaluation program that utilizes student performance data accounted for significant differences within groups. ANOVA tests were run using SPSS with the significance level set at p < .05. Table 37 documents how teachers’ perceptions did significantly differ based on the teacher’s experience with an evaluation program that utilizes student performance data, F (2,163) = 20.947, p = 0.001.
Tukey post-hoc analysis revealed that this difference was attributable to teachers with one year or more experience with an evaluation program that utilizes student performance data responding more favorable to accuracy standard questions than did teachers in their first year with such an evaluation program and teachers with no experience. There was also a significant difference in perceptions of teachers in their first year with an evaluation program that utilizes student performance data than teachers with no experience. Table 38 shows the post-hoc results.

Table 38: Tukey Post-hoc Analysis for Accuracy Standard by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data

<table>
<thead>
<tr>
<th>(I) Exp</th>
<th>(J) Exp</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>1 Year</td>
<td>First</td>
<td>-.40208*</td>
<td>.13515</td>
<td>.009</td>
<td>-.7217</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>-.66160*</td>
<td>.11793</td>
<td>.001</td>
<td>-.9405</td>
</tr>
<tr>
<td>First</td>
<td>1 Year</td>
<td>.40208*</td>
<td>.13515</td>
<td>.009</td>
<td>.0824</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>-.25952*</td>
<td>.10947</td>
<td>.049</td>
<td>-.5185</td>
</tr>
<tr>
<td>None</td>
<td>1 Year</td>
<td>.66160*</td>
<td>.11793</td>
<td>.001</td>
<td>.3827</td>
</tr>
<tr>
<td></td>
<td>First</td>
<td>.25952*</td>
<td>.10947</td>
<td>.049</td>
<td>.0006</td>
</tr>
</tbody>
</table>

*. = The mean difference is significant as the 0.05 level.

Teachers were provided the opportunity to list up to three benefits and limitations associated with including student performance data in their evaluation. The teacher's feedback related to the accuracy standard on the potential benefits from including student performance data in evaluations indicated that student performance data would "translate to a more meaningful and effective evaluation for once" and "finally allow for effective teachers (who are
not necessarily the ‘favorites’) to be recognized.” The teachers’ feedback on the limitations of including student performance data associated with the accuracy standard noted that current standardized tests are “invalid,” “unreliable,” and “bad indicators of student progress.” Teachers also commented on how students “don’t take the tests seriously” and “punish their teachers by bombing the end of year tests.” The open-ended responses were coded and grouped into similar constructs. The potential benefits and limitations cited by teachers associated with the accuracy standard are listed in Table 39.

Table 39: Teacher-cited Benefits and Limitations Associated with the Accuracy Standard

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Makes Evaluation More Objective (3)</td>
<td>Inaccurate Assessments (5)</td>
</tr>
<tr>
<td>Identifies Good Teachers</td>
<td>Student Apathy Toward Test (10)</td>
</tr>
<tr>
<td></td>
<td>Does Not Account for Student Ability Groupings (5)</td>
</tr>
<tr>
<td></td>
<td>Teaching Students with Disabilities or English Language Learners (3)</td>
</tr>
</tbody>
</table>

Research Question Five

Is there a significant difference in the perceptions of teachers toward the incorporation of student performance data in their evaluation among teachers (a) with different years of experience; (b) in union and non-union states; (c) in tested and non-tested grades and courses; and (d) in elementary, middle, and high schools?

Years of Experience. Data were analyzed to determine whether specific demographic criteria accounted for significant differences within groups. ANOVA tests were run using SPSS with the significance level set at $p < .05$. To determine the effect of demographic criteria, the demographic criteria served as the independent variable while the dependent variable alternated between the four evaluation standards (propriety, utility, feasibility, and accuracy). In determining the impact of the teachers’ years of experience, the teachers’ years of experience
served as the constant independent variable while the dependent variable alternated between each of the four evaluation standards. Table 40 documents how teachers’ perceptions did not significantly differ based on the teacher’s years of experience in any of the four evaluation standards.

Table 40: Teacher ANOVA by Years of Experience

<table>
<thead>
<tr>
<th>Evaluation Standard</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propriety Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>1.079</td>
<td>2</td>
<td>.539</td>
<td>1.193</td>
<td>.306</td>
</tr>
<tr>
<td>Within Groups</td>
<td>73.7</td>
<td>163</td>
<td>.452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>74.78</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.496</td>
<td>2</td>
<td>.248</td>
<td>1.111</td>
<td>.332</td>
</tr>
<tr>
<td>Within Groups</td>
<td>36.39</td>
<td>163</td>
<td>.223</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36.89</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasibility Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.868</td>
<td>2</td>
<td>.434</td>
<td>1.489</td>
<td>.229</td>
</tr>
<tr>
<td>Within Groups</td>
<td>47.5</td>
<td>163</td>
<td>.291</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48.36</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.526</td>
<td>2</td>
<td>.263</td>
<td>1.163</td>
<td>.315</td>
</tr>
<tr>
<td>Within Groups</td>
<td>36.87</td>
<td>163</td>
<td>.226</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>37.39</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Union and Non-Union States. Whether teachers were working in a school district that operated under a collective bargaining agreement or in a school district without a collective bargaining agreement did not produce a meaningful difference in their responses. Table 41 documents the results for the ANOVA run using the data set by the teacher’s years of experience.
### Table 41: Teacher ANOVA by Participation under a Collective Bargaining Agreement

<table>
<thead>
<tr>
<th>Evaluation Standard</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propriety Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.114</td>
<td>1</td>
<td>.114</td>
<td>.296</td>
<td>.587</td>
</tr>
<tr>
<td>Within Groups</td>
<td>63.29</td>
<td>164</td>
<td>.386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>63.4</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.189</td>
<td>1</td>
<td>.189</td>
<td>.933</td>
<td>.336</td>
</tr>
<tr>
<td>Within Groups</td>
<td>33.2</td>
<td>164</td>
<td>.202</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>33.39</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasibility Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.022</td>
<td>1</td>
<td>.022</td>
<td>.076</td>
<td>.783</td>
</tr>
<tr>
<td>Within Groups</td>
<td>48.34</td>
<td>164</td>
<td>.295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48.36</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>0.065</td>
<td>1</td>
<td>.065</td>
<td>.248</td>
<td>.619</td>
</tr>
<tr>
<td>Within Groups</td>
<td>43.11</td>
<td>164</td>
<td>.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>43.18</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tested and non-tested grades and courses.** Whether teachers were working in a school district that operated under a collective bargaining agreement or in a school district without a collective bargaining agreement did not produce a meaningful difference in their responses.

Table 42 documents the results for the ANOVA run using the data set by the teacher’s years of experience.
Table 42: Teacher ANOVA by Tested and Non-tested Grades and Courses

<table>
<thead>
<tr>
<th>Evaluation Standard</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propriety Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.295</td>
<td>1</td>
<td>.295</td>
<td>.751</td>
<td>.388</td>
</tr>
<tr>
<td>Within Groups</td>
<td>64.51</td>
<td>164</td>
<td>.393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>64.81</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.152</td>
<td>1</td>
<td>.152</td>
<td>.751</td>
<td>.388</td>
</tr>
<tr>
<td>Within Groups</td>
<td>33.23</td>
<td>164</td>
<td>.203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>33.39</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasibility Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.192</td>
<td>1</td>
<td>.192</td>
<td>.654</td>
<td>.420</td>
</tr>
<tr>
<td>Within Groups</td>
<td>48.17</td>
<td>164</td>
<td>.294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48.36</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy Standard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.19</td>
<td>1</td>
<td>.194</td>
<td>.769</td>
<td>.382</td>
</tr>
<tr>
<td>Within Groups</td>
<td>41.27</td>
<td>164</td>
<td>.252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>41.47</td>
<td>165</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Level of School. Whether teachers worked in an elementary, middle, or high school did not significantly impact teacher perceptions to the use of student performance data in their evaluations. Table 43 identifies the results for the ANOVA run using the four evaluation standards as the dependent variables and the level of school as the independent variable.
Table 43: Teacher ANOVA by Level of School

<table>
<thead>
<tr>
<th>Evaluation Standard</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propriety Standard</td>
<td>Between Groups</td>
<td>.022</td>
<td>2</td>
<td>.011</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>74.76</td>
<td>163</td>
<td>.459</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>74.78</td>
<td>165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility Standard</td>
<td>Between Groups</td>
<td>.173</td>
<td>2</td>
<td>.086</td>
<td>.423</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>33.21</td>
<td>163</td>
<td>.204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>33.39</td>
<td>165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasibility Standard</td>
<td>Between Groups</td>
<td>.561</td>
<td>2</td>
<td>.281</td>
<td>.957</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>47.8</td>
<td>163</td>
<td>.293</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>48.36</td>
<td>165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy Standard</td>
<td>Between Groups</td>
<td>.089</td>
<td>2</td>
<td>.044</td>
<td>.177</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>40.84</td>
<td>163</td>
<td>.251</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>40.93</td>
<td>165</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cross-Comparative Analysis of Teacher Perceptions Based on Level of Experience with Evaluation that Utilizes Student Performance

The only demographic category that produced significant differences in perceptions was how much experience the teacher had with working with an evaluation program that utilized student performance data. Teachers in their first year of such an evaluation program and teachers with one year or more experience with such a program were more favorable to how the inclusion of student performance data positively impacted the propriety, utility, and accuracy standards. ANOVA tables presented earlier in this chapter demonstrate that these differences among groups were statistically significant (p < .01) for all three standards. Regardless of their experience with an evaluation program that utilizes student performance data, teachers were similar in their responses to questions associated with the feasibility standard. Table 44 provides
a comprehensive view for all four evaluation standards. A value of 1 indicates strong agreement, 2 indicates agreement, 3 indicates disagreement and a value of 4 indicates strong disagreement.

Table 44: Teacher Mean & Standard Deviation (SD) for Evaluation Standards Disaggregated by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data

<table>
<thead>
<tr>
<th>Level of Experience</th>
<th>Teacher Sample N = 166</th>
<th>Propriety Standard Mean</th>
<th>Propriety Standard SD</th>
<th>Utility Standard Mean</th>
<th>Utility Standard SD</th>
<th>Feasibility Standard Mean</th>
<th>Feasibility Standard SD</th>
<th>Accuracy Standard Mean</th>
<th>Accuracy Standard SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Experience</td>
<td>90</td>
<td>3.14</td>
<td>.184</td>
<td>2.61</td>
<td>.294</td>
<td>2.47</td>
<td>.341</td>
<td>3.06</td>
<td>.259</td>
</tr>
<tr>
<td>First Year of Program</td>
<td>42</td>
<td>2.91</td>
<td>.446</td>
<td>2.28</td>
<td>.209</td>
<td>2.43</td>
<td>.302</td>
<td>2.80</td>
<td>.233</td>
</tr>
<tr>
<td>1 Year or More</td>
<td>34</td>
<td>2.49</td>
<td>.233</td>
<td>2.09</td>
<td>.289</td>
<td>2.42</td>
<td>.342</td>
<td>2.40</td>
<td>.306</td>
</tr>
</tbody>
</table>

Chart 1 provides a visual representation of the teacher perception data with respect to teacher experience with an evaluation program that utilizes student performance data. It documents how closely all participants responded to survey questions associated with the feasibility standard. The chart also highlights how teachers with one year or more experience consistently responded the most favorable to questions associated with the other three evaluation standards.
Chart 1: Teacher Mean for Evaluation Standards Disaggregated by Teacher Experience with an Evaluation Program that Utilizes Student Performance Data
Chapter 5: Summary and Discussion of Findings

Teachers deserve the opportunity to be evaluated using meaningful and objective data. Maintaining conventional teacher evaluation programs that do not take into account student achievement data jeopardizes opportunities for growth for teachers as well as students. There is substantial research validating the impact an effective teacher has on student achievement (see, for example, Stronge, 2010; Stronge, Ward, & Grant, 2011). In an era where valid and reliable student performance data are becoming more readily available, school leaders and policymakers must now use this valuable source of data as a component in evaluating teaching and learning.

Teacher evaluations possess the ability to spotlight strengths in a teacher’s delivery of instruction and identify where professional growth and development may be necessary. This opportunity to distinguish and improve teaching and learning can only be fully realized when evaluations accurately reflect the intended outcome measures—student performance data. There is also research that documents the importance of teacher buy-in on the success of new educational initiations (Bryk et al., 1999; Greene & Lee, 2006; Schneider & Bryk, 2000; Turnbull, 2002). It is due to these reasons that teacher evaluations should include student performance data and why the perceptions of teachers in the use of student performance data in teacher evaluations also must be considered.

The motivation for the increased level of attention toward teacher evaluations has been, in large part, due to a new era of accountability ushered in by the No Child Left Behind (NCLB) Act of 2001 and subsequent legislation. In 2010, the United States Department of Education specifically called for improving teacher effectiveness to ensure that every classroom has a great teacher in it (United States Department of Education, 2010). As part of this initiative, teacher evaluations must focus on recognizing, encouraging, and rewarding excellence and allow for
teachers to receive meaningful information about their practice (United States Department of Education, 2010, p. 4). The Joint Committee on Standards for Educational Evaluation (JCSEE) (2009) constructed a refined set of personnel evaluation standards to help respond to this finding. This committee created and has since expanded a set of standards that address four accepted attributes of educational evaluation: propriety, utility, feasibility, and accuracy (2009).

Subsequent research affirmed that these four attributes cited by the JCSEE are required tenants in any sound teacher evaluation (Howard & Gullickson, 2010).

A wealth of research demonstrates that the single most important factor in a student’s level of academic achievement is predicated by the effectiveness of the student’s teacher (Aaronson et al., 2007; Goldhaber & Hansen, 2008; Hattie, 2009; Heck, 2009; Marzano, 2003a; Nye, Konstantopolulos, & Hedges, 2004; Palardy & Rumberger, 2008; Rivkin et al., 2005; Rothstein, 2010; Sass, 2008; Stronge, 2010; Stronge, Ward, & Grant, 2011). Research clearly documents the strong correlation an effective teacher has on a student’s achievement gains during the school year as measured by value-added or other test-based growth measures (Goe et al., 2008; Wright, Horn, & Sanders, 1997). Since the research clearly demonstrates that the quality of teaching matters, it is reasonable to presume that a quality teacher evaluation process also matters in order to know if the school system possesses high quality teachers (Stronge & Tucker, 2003). As a result of these various initiatives, State Departments of Education and school districts began developing evaluation matrices that attempt to leverage the power of student performance data to complement other teacher performance domains to more effectively evaluate teachers.

This study sought to identify K-12 public school teachers’ perceptions regarding the use of student performance data in teacher evaluation. Specifically, this study sought to discover to
what extent teachers felt the use of student performance data in teacher evaluations impacted the propriety, utility, feasibility, and accuracy standards of the evaluation program. In order for states and school districts to realize the expected goals from including student performance data to teacher evaluations, it is imperative for instructional leaders to understand how teachers perceive this change. Although teacher buy-in or support is not required for the changes to the teacher evaluation process, there is research that documents that educational reform programs with teacher support have greater opportunities for lasting success (Bryk et al., 1999; Greene & Lee, 2006; Schneider & Bryk, 2000; Turnbull, 2002). Once teachers understand how student performance data have the opportunity to complement other performance domains in their evaluation, the enhanced teacher evaluation program may gain sustained support from all stakeholders.

Data were collected from a national stratified random sample of 166 K-12 public school teachers who completed an online survey in February and March of 2013. Teachers were asked to respond to statements about the use of student performance data in their evaluation. The researcher specifically sought to determine how teachers perceived the use of student performance data in teacher evaluations with respect to the four evaluation standards presented by the Joint Committee on Standards for Educational Evaluation (JCSEE) and if certain demographic characteristics impacted those perceptions. Results were disaggregated by participants’ responses to demographic factors and analyzed for statistical significance. Demographic information was solicited in the final four items of the survey. That information included: (a) number of years experience in education; (b) whether the teacher taught under a collective bargaining agreement; (c) whether the teacher taught a tested or non-tested grade or course; and (d) level of school in which the teacher worked (elementary, middle, or high).
The purpose of the current study was to determine how teachers perceived the use of student performance data in their evaluation and to determine if demographic features significantly influenced those perceptions.

Results documented a general agreement among teachers with respect to how they perceived the use of student performance data in their evaluations. This level of agreement spanned across all four JCCSS evaluation standards. Additionally, there were no significant differences in teachers’ responses based on the various demographic factors. The one category that did produce significant differences in teachers’ responses was discovered when disaggregating the teacher’s experience with an evaluation program that included student performance data. Teachers with more than one year of experience in this type of evaluation program were more likely to respond favorably toward the use of student performance data in their evaluation in the propriety, utility, and accuracy evaluation standards. The difference in perceptions in the propriety and accuracy standards was especially large. Participants in the survey were asked to provide additional items that they viewed as a benefit or limitation to the use of student performance data in teacher evaluations. Content analysis documented that teachers identified very few additional considerations from having student performance data included in teacher evaluations. The only new considerations identified by teachers as benefits included: increases teacher collaboration, increases focus on test-taking skills, and better assists common core alignment. New limitations associated with using student performance data identified by teachers included: concerns about teachers who work with students with disabilities or English Language Learners, student apathy toward the assessment, and fear that the student data would become the sole source of the teacher’s evaluation. The higher frequency of negative comments indicated that teachers opposed to the utilization of student performance data were
more inclined to voice these feelings in more detail in this section of the survey. The vast majority of responses provided by teachers repeated topics and constructs already included in the survey instrument (see Appendix E).

Discussion of Findings Related to Research Questions

Discussion Related to Research Question One

*To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the propriety standard of the Joint Committee Personnel Evaluation Standards?*

Teachers participating in this study were asked to respond to four questions that were associated with the propriety standard of the Joint Committee Personnel Evaluation Standards. The teachers were asked whether they Strongly Agree, Agree, Disagree, or Strongly Disagree with the statement. A rank of 1 indicated that the teacher marked Strongly Agree, a rank of 2 indicated the teacher selected Agree, a rank of 3 indicated the teacher chose Disagree, and a rank of 4 indicated that the teacher selected Strongly Disagree. Descriptive statistics were calculated for each statement, including mean and standard deviation. The mean for each of the four statements were 2.95, 2.95, 2.97, and 2.93 producing a range of 0.04. The standard deviation ranged from 0.697 to 0.708 for the four statements. The reliability coefficient for the questions was 0.787 indicating acceptable reliability.

The propriety standard demonstrates whether the rights of the individuals affected by an evaluation are protected. It specifically determines whether the evaluation system is conducted ethically, legally, and with regard for the personal welfare of the individuals involved in the evaluation (Joint Committee on Standards for Educational Evaluation, 2009). The four questions in this survey specifically attempted to identify whether student performance data improves the
evaluation process, protects the rights of teachers, and enhances the ability to provide ratings that differentiate between levels of performance. The results in Table 17 indicate that teachers did not perceive the use of student performance data as positively impacting the propriety standard in their evaluation.

The teachers' responses to the open-ended questions provided additional feedback on how they perceived evaluations that used student performance data. Teachers were asked to provide examples of the benefits and limitations regarding using student performance data in evaluations (see Appendix E). With respect to the propriety standard, teachers noted that using student performance data would benefit evaluations because this data "identifies good teachers" and improves the overall evaluation process by making the program "more objective" and "removing evaluator bias." Fourteen teachers feared that student performance data would become the "sole source of evaluation data" which would not be especially fair to teachers in schools that have historically poor academic results or teach students who have historically scored low on assessments. This perceived fear was the most often reported limitation cited by teachers in the open-ended section. The responses in the open-ended section of the survey were evenly divided among all demographic groups.

An evaluation process adhering to propriety standards that more distinctly differentiates between levels of performance likely faces considerable obstacles from teachers who have routinely been rated as satisfactory and above for decades (Weisberg et al., 2009). It is, therefore, not entirely surprising to see the results in this study reflect a reluctance to embrace a new evaluation program that now includes student performance data. The fact that teachers who had never participated in an evaluation program with student performance data overwhelmingly chose "Disagree" and "Strongly Disagree" over other choices to the four questions associated
with propriety standards demonstrates this heightened level of concern over the use of student performance data in evaluations (see Chart 2).

Teachers who had reported experience of working in a school district using an evaluation process that used student performance data were less likely to disagree with these four questions than those teachers who had never participated in such an evaluation process. The 34 teachers who had more than one year of experience had a mean score of 2.29 (indicating agreement) versus a mean score of 3.22 (indicating disagreement) for the 90 teachers without any experience with an evaluation program that utilized student performance data. The 42 teachers in the survey who were in their first year of such an evaluation program had a mean score of 2.90 which was similar to the overall mean. For the four propriety questions in this survey, Chart 2 illustrates the percentages of each response to the propriety standard questions broken down by amount of experience the teacher had with student performance data (SPD) in his or her evaluation program. The y-axis in Chart 2 notes the percentage by which each subgroup responded with Strongly Agree, Agree, Disagree, and Strongly Disagree.

The heightened level of disagreement from the 90 teachers without experience with an evaluation program that uses student performance data suggests that the fear of the unknown associated with student performance data as a component in an evaluation program may contribute to the overall level of disagreement in this standard. Research into what teachers feared from evaluations confirms this supposition. Emery and Ohanian (2004, p. 34) reported that teachers were fearful of what harm or consequences would come to them as a result of test results interpreted incorrectly by principals or district officials. Teachers also expressed concerns that this level of increased accountability placed so much pressure on teachers that many of them would resort to "teaching to the test" (Knight, 2008). These specific fears were
noted in the open-ended section by teachers. Teachers specifically commented that “data now will replace everything else I do at the school which can’t be quantified in numbers,” “test scores will trump all in the evaluation,” and “my principal can’t understand scores and I am afraid it will hurt me.”

Chart 2: Responses to Propriety Questions by Participants’ Experience with Student Performance Data (SPD) in Evaluation Program

The fact that teachers with experience in an evaluation program that uses student performance data are more positive in their survey responses is consistent with other studies. In a number of studies, teachers report that they desire knowing what standards or indicators they will be evaluated against and how this evaluation will be conducted (Conley, Muncey & You, 2005; Giliya, 2006; Milanowski & Heneman, 2001; Pizzi, 2009; Sand, 2005; Sanders, 2000; Seyfarth, 2002). Teachers also report that they favor more transparent evaluation programs (Castillo, 2005; Feeney, 2007; Sand, 2005). Teachers specifically comment that they want to know how they will be evaluated and how the evaluation program will be conducted (Castillo 2005; Giliya, 2006; Musick, 1997; Pizzi, 2009; Sand, 2005). The introduction of student performance data into the evaluation program satisfies all of these desires from teachers. It is
possible that a similar study conducted after teachers have more experience and knowledge of how an evaluation program that uses student performance data will produce more favorable responses toward questions associated with propriety standards in this study.

Discussion Related to Research Question Two

To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the utility standard of the Joint Committee Personnel Evaluation Standards?

There were ten questions associated with the utility standard of the Joint Committee Personnel Evaluation Standards in the survey completed by teachers. The 166 teachers were asked whether they Strongly Agree, Agree, Disagree, or Strongly Disagree with the statement. A rank of 1 indicated that the teacher marked Strongly Agree, a rank of 2 indicated the teacher selected Agree, a rank of 3 indicated the teacher chose Disagree, and a rank of 4 indicated that the teacher marked Strongly Disagree. Descriptive statistics were calculated for each statement, including mean and standard deviation. The range of means for the four statements was 0.16 (2.52-2.36). The standard deviation ranged from 0.757 to 0.947 for the ten statements. The reliability coefficient for the questions was 0.751 indicating acceptable reliability.

The utility standards examine whether evaluations are timely, informative, and influential. In particular, evaluation systems that adhere to the utility standards include a constructive orientation and provide useful information which assists in the improvement of individual and group performance (Joint Committee on Standards for Educational Evaluation, 2009). The ten questions in this survey attempted to identify whether student performance data in the evaluation process informs and improves teaching, increases attention on the use of assessment data that improves teaching performance, appropriately evaluates the job
expectations of teachers, and informs professional development. The results in Table 23 indicate that teachers were more favorable to how student performance data would improve the evaluation process with respect to the utility standard as compared to the other three evaluation standards.

The teachers’ responses to the open-ended questions reaffirmed the teachers’ responses to the four-point Likert scale questions (see Appendix E). Teachers cited more examples of how including student performance data in evaluation programs would benefit the utility nature of the evaluation than any of the other three standards. Teachers commented that an evaluation program that uses student performance data would “guide lesson planning,” “identify student gaps,” “inform professional development,” and “enhance personal growth and reflection.” The responses also identified some perceived liabilities. The most prominent fear was provided by 16 teachers who feared that the inclusion of student performance data would be used to promote “teaching to the test.” The comments associated with the utility standard were fairly evenly divided among all demographic groups.

The teachers’ perceptions are aligned to research that suggested most evaluation programs did little to improve practice or instruction (Peterson, 2000) and can become “little more than a time-consuming charade” (Stronge & Tucker, 2003, p. 6). Previous studies report that teachers do not perceive current evaluation programs as substantially improving their teaching which would fall under the utility standard. Teachers and administrators each perform their assigned role in the evaluation process and not surprisingly very few substantial changes in teaching and learning transpired (Weisberg et al., 2009). Teachers did not regard evaluations, based on two or three formal observations and using a checklist to determine observed strategies, instrumental in improving teaching (Colby et al., 2002; Levandowski, 2000; Sand, 2005; Sutton,
This perception is supported by a large-scale study by Kauchak, Peterson, and Driscoll (1985) when they surveyed teachers in Utah and Florida. Teachers in this study reported that their current evaluation program had minimal impact on their teaching.

Teachers in the study who have participated in an evaluation program that uses student performance data specifically reported that such a program presents teachers with more value since student achievement data could better guide professional growth and development. As noted with the propriety standard, teachers who had previous experience with an evaluation process that included student performance data were prone to provide more favorable responses in this section. The 42 teachers in their first year of an evaluation program and the 34 teachers with more than one year using student performance data had a combined mean of 2.08 (indicating agreement) while the 90 teachers with no experience had a mean of 2.74 (indicating disagreement). Chart 3 illustrates the percentages of each response to the utility standard questions broken down by amount of experience the teacher had with student performance data (SPD) in his or her evaluation program. The y-axis in Chart 3 notes the percentage by which each subgroup responded with Strongly Agree, Agree, Disagree, and Strongly Disagree. This finding suggests that teachers with increasing exposure to an evaluation program that includes student performance data are more likely to agree to statements that state student performance data assists the utility nature of teacher evaluations.
Research Related to Research Question Three

To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the feasibility standard of the Joint Committee Personnel Evaluation Standards?

There were four questions associated with the feasibility standard of the Joint Committee Personnel Evaluation Standards in the survey completed by teachers. The 166 teachers were asked whether they Strongly Agree, Agree, Disagree, or Strongly Disagree with the statement. A rank of 1 indicated that the teacher marked Strongly Agree, a rank of 2 indicated the teacher selected Agree, a rank of 3 indicated the teacher chose Disagree, and a rank of 4 indicated that the teacher marked Strongly Disagree. Descriptive statistics were calculated for each statement, including mean and standard deviation. The mean for each of the four statements were 2.47, 2.43, 2.46, and 2.45 producing a range of 0.04. The standard deviation ranged from 0.624 to...
0.893 for the four statements. The reliability coefficient for the four questions was 0.774 indicating acceptable reliability.

The feasibility standards relate to whether the evaluation system is relatively easy to implement, efficient in the use of time and resources, adequately funded and politically viable (Joint Committee on Standards for Educational Evaluation, 2009). The four questions in this survey attempted to identify whether student performance data in the evaluation process can effectively be used to evaluate a teacher and whether this information can describe the context surrounding the data collected. The results in Table 29 indicate that teachers were divided on how the use of student performance data impacted the feasibility standard in their evaluation. The teachers' responses to the open-ended questions did not specifically reference any topics associated with the feasibility standard. The amount of experience a teacher had with an evaluation program that used student performance data did not significantly impact the teacher's response in these four questions. The small number of questions associated with the feasibility standard also may have prevented more substantial findings.

Research Related to Research Question Four

To what degree do teachers perceive the use of student performance data in their teacher evaluation as aligned with the accuracy standard of the Joint Committee Personnel Evaluation Standards?

Teachers participating in this study were asked to respond to twelve questions that were associated with the accuracy standard of the Joint Committee Personnel Evaluation Standards. The accuracy standard accounted for the largest number of questions in the survey. The teachers were asked whether they Strongly Agree, Agree, Disagree, or Strongly Disagree with the statement. A rank of 1 indicated that the teacher marked Strongly Agree, a rank of 2 indicated
the teacher selected Agree, a rank of 3 indicated the teacher chose Disagree, and a rank of 4 indicated that the teacher marked Strongly Disagree. Descriptive statistics were calculated for each statement, including mean and standard deviation. The mean for the twelve statements ranged from a high of 2.90 to a low of 2.80. The standard deviation ranged from 0.726 to 0.813 for the twelve statements. The reliability coefficient for the questions was 0.868 indicating good reliability.

The accuracy standard demonstrates whether “personnel evaluations allow evaluators to make sound judgments and decisions; whether the evaluation methodology is appropriate for the purpose of the evaluation, the individuals being evaluated, and their work contest; and whether the evaluation are valid and reliable” (Joint Committee on Standards for Educational Evaluation, 2009, p. 7). The twelve questions in this survey specifically attempted to identify whether teachers perceive student performance data improving the accuracy of the evaluation results for them and for their colleagues and whether the evaluation is based on justifiable and documented performance. The results in Table 23 indicate that teachers did not perceive the use of student performance data as positively impacting the accuracy standard in their evaluation. The response “Disagree” was the most often selected by teachers for these four statements (see Chart 4).

The teachers’ responses to the open-ended questions reiterated many of the teachers’ responses to the four-point Likert scale questions (see Appendix E). Five teachers noted in their responses that they feared “inaccurate assessments” would translate to inaccurate evaluation scores. Seven teachers commented that the “misuse of data” by their administrator would further damage the accuracy of their evaluations. The benefits cited by teachers included comments that student performance data would make evaluations more “objective” and remove “evaluator bias.” The open-ended responses were fairly well dispersed among all demographic categories.
The survey responses taken as a whole appear to contradict other research in this area. For example, teachers have reported that the substance of current evaluation programs is often marginalized because it fails to provide an accurate report of what transpires in the classroom on a regular basis (Castillo, 2005; Colby et al., 2002; Giliya, 2006; Levandowski, 2000). Teachers also critique evaluation programs for the subjectivity found in summative administrator evaluation reports and the administrator’s lack of competence and resolve to evaluate accurately (Castillo, 2005; Cooper et al., 2005; Sand, 2005). The lack of time an administrator has prevents him or her to evaluate every teacher accurately or follow up with the teacher appropriately (Jacob & Lefgren, 2008; Toch & Rothman, 2008; Weisberg, Sexton, Mulher, & Keeling, 2009). With the large number of studies documenting how teachers feel their evaluation is not an accurate representation of their teaching or their colleagues’ teaching, it is interesting to discover that teachers are reluctant to want student performance data to be used to potentially increase the accuracy of these evaluations.

Once again, the researcher was able to identify some interesting conclusions by disaggregating the survey results by amount of experience with an evaluation program that uses student performance data. Again, teachers with more than one year of experience with an evaluation program that uses student performance data were more apt to agree with statements that student performance data provides more accurate evaluations. The mean for these teachers with experience to the twelve accuracy questions was 2.24 (indicating agreement). The mean for teachers in their first year of an evaluation program with student performance data rose to 2.79 and the mean for teachers without any experience increased to 3.10 (indicating disagreement). This finding demonstrates that as teachers are more familiar with an evaluation program that uses student performance data they are more likely to report that such a program improves the overall
accuracy of the evaluation. Chart 4 illustrates the percentages of each response to the accuracy standard questions broken down by amount of experience the teacher had with student performance data (SPD) in his or her evaluation program. The y-axis in Chart 4 notes the percentage by which each subgroup responded with Strongly Agree, Agree, Disagree, and Strongly Disagree.

Chart 4: Responses to Accuracy Questions by Participants' Experience with Student Performance Data (SPD) in Evaluation Program

Research Related to Research Question Five

Is there a significant difference in the perceptions of teachers toward the incorporation of student performance data in their evaluation among teachers (a) with different years of experience; (b) in union and non-union states; (c) in tested and non-tested grades and courses; and (d) in elementary, middle, and high schools?

Teachers in this study responded to questions about their perceptions to the use of student performance data in teacher evaluations. The questions were then assigned to the personnel evaluation standards created by the Joint Committee on Standards for Educational Evaluation.
These standards are organized into four inter-related areas: propriety, utility, feasibility, and accuracy. Each standard was tested for statistical significance using one-way ANOVAs. For each ANOVA one of the following demographic factors served as the independent variable: (a) years of experience, (b) union or non-union state; (c) tested or non-tested grade and course; and (d) elementary, middle, or high school. Results of the study revealed that no significant differences were found.

A review of the findings from the study follows. The review is organized by demographic characteristics of the teachers and the schools where they teach—years of teaching experience, presence of a collective bargaining agreement, teaching a tested or non-tested grade and course, and level of school. The purpose of this framework provides additional insight into how much each demographic factor played influenced the teachers' perceptions about the use of student performance data in teacher evaluations.

**Years of Experience.** A demographic question posed to every teacher completing the survey asked them to identify their years of teaching experience. Teachers selected one of three possible ranges to represent their total years of teaching experience: (a) 0-4 years, (b) 5-10 years, and (c) 11 or more years of teaching experience. Over half of the teachers had over 10 years of teaching experience (54%). Teachers with 0-4 years accounted for 20% of the survey respondents and teachers with 5-10 years of experience accounted for the remaining 26% of participants. Results of the study indicate that the years of teaching experience did not account for any significant differences among responses in any of the four evaluation standards. This suggests that teacher perceptions regarding the use of student performance data in teacher evaluations is not influenced by the teachers' teaching experience. Caution should be used when making inferences about this finding because of the low number of participants in the study.
Although years of experience accounted for no significant differences, the researcher sought to draw some meaningful conclusions from any patterns that emerged from teacher responses on the same evaluation standard for the same variable (in this case, years of experience). No discernible pattern emerged among these differences.

**Union or non-union.** The second demographic question included in the survey asked every teacher completing the survey whether they worked under a collective bargaining agreement. Teachers either selected yes or no to this question. Teachers were fairly closely split in this demographic with 54% of teachers indicating they currently work under a collective bargaining agreement and 46% of teachers noting that they do not work under a collective bargaining agreement. Results of the study indicate that working under a collective bargaining agreement did not account for any significant differences among responses in any of the four evaluation standards. This suggests that teacher perceptions regarding the use of student performance data in teacher evaluations are not influenced by whether a teacher works under a collective bargaining agreement. Caution should be used when making inferences about this finding because of the low number of participants in the study.

Even though teaching under a collective bargaining agreement did not yield any significant differences, the researcher sought to draw some meaningful conclusions from any patterns that emerged from teacher responses on the same evaluation standard for the same variable (in this case, working or not working under a collective bargaining agreement). No discernible pattern emerged among these differences.

**Tested or non-tested grade and course.** Whether the teacher taught a tested or non-tested grade and course represented the third demographic question posed to every teacher completing the survey. Teachers either selected yes or no to this question. Teachers who taught
a tested grade and course represented 41% of the survey respondents while the remaining 59% of respondents answered that they do not teach a tested grade or course. Results of the study indicate that the whether a teacher taught a tested or non-tested grade or course did not account for any significant differences among responses in any of the four evaluation standards. This suggests that teacher perceptions regarding the use of student performance data in teacher evaluations is not influenced by the tested or non-tested nature of the grade or course the teacher teaches. Caution should be used when making inferences about this finding because of the low number of participants in the study.

The researcher sought to draw some meaningful conclusions from any patterns that emerged from teacher responses on the same evaluation standard for the same variable (in this case, tested or non-tested grade or course). No discernible pattern emerged among these differences.

Level of school. A fourth demographic question posed to every teacher completing the survey asked them to identify the level of school where they taught. Teachers selected one of three possible ranges to represent their level of school: (a) elementary school, (b) middle school, and (c) high school. Just over half of the teachers taught at the elementary school level (55%). Middle school teachers accounted for 17% of the teachers in the survey and high school teachers accounted for the remaining 28% of teachers. Results of the study indicate that the level of school where a teacher taught did not account for any significant differences among responses in any of the four evaluation standards. This suggests that teacher perceptions regarding the use of student performance data in teacher evaluations is not influenced by the level of school where the teacher teaches. Caution should be used when making inferences about this finding because of the low number of participants in the study.
Although level of school accounted for no significant differences in the ANOVA test, the researcher sought to draw some meaningful conclusions from any patterns that emerged from teacher responses on the same evaluation standard for the same variable (in this case, level of school). No discernible pattern emerged among these differences.

**General Discussion of Findings**

**Validation Support for Survey Instrument**

This study provides evidence that the survey used is a valid and reliable instrument to identify teacher perceptions of the use of student performance data in teacher evaluations. The survey was developed for this study based initially on the research and work conducted by Joan Herman and Shari Golan on teachers’ perceptions of standardized testing and its impact on teachers and learners (1991) and Ansie Lessing on teachers’ perceptions of the value of professional development (2007). Herman and Golan’s survey instrument was adapted with written permission from the researchers and through the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) by The Regents of the University of California as supported under the Institute of Education Services (IES), U.S. Department of Education. Lessing’s survey was adapted with written permission from the author. The adapted survey instrument was then field tested with a group of doctoral students. Twelve doctoral students at The College of William & Mary with varying levels of teaching, administrative, and other education-related experience participated in the first field test.

The survey was then reviewed by an expert panel. For the purposes of this study, an expert is defined as an individual with extraordinary insight into the population and/or subject beyond what a member of the population under study or participant in the phenomenon being investigated might have (Ramirez, 2002). This four-member expert panel included individuals
who all possessed a doctoral degree in Educational Policy and Leadership and considerable experience in the design, implementation, and review of scholarly research. Dr. Min Sun, Assistant Professor in Educational Policy and Quantitative Methods at Virginia Polytechnic Institute and State University, Dr. Leslie Grant, Assistant Professor of Education at The College of William & Mary, Dr. Marco Munoz, Evaluation Specialist in the Data Management, Planning, and Program Evaluation Services Division at the Jefferson County Public Schools (Louisville, Kentucky), and Dr. Virginia Tonneson, educational consultant and recent contributor to Virginia Department of Education State-wide Teacher Evaluation Project, served as reviewers. Research in the development of valid and reliable surveys documents that expert reviewers have the ability to improve in surveys by providing input on the content of the questionnaire, importance and meaningfulness of question areas to research aims, and wording and terminology of items (Dillman, 2002). A final expert panel composed of Dr. Jennifer Hindman, Coordinator at the School Leadership Institute and the School University Research Network at The College of William & Mary, Amy Colley, Assistant Superintendent of Instruction and Support Services with Poquoson, Virginia, Public Schools, and Dr. Lisa Pennycuff, Director of Accountability and Instructional Services with York County, Virginia, Public Schools, reviewed the final survey to create a table of specifications that organized questions into each of the four JCSEE evaluation standards.

In both the field test and the expert panel reviews of the survey, participants reviewed the statements, directions, and format of the survey. These groups were also testing the survey to ensure that the statements in the survey included content relevant to the study and research questions. This expert review additionally helped to determine the credibility, conformability,
and dependability of the survey instrument. Input from the field test and expert review was used
to create the final survey instrument.

The reliability of the survey instrument was confirmed after a comprehensive analysis of
the survey responses. The internal consistency of responses was addressed using Cronbach’s
alpha. Table 45 demonstrates how this measure of internal consistency determined that
questions within each of the four JCSEE evaluation standards were closely related at the
acceptable and good ranges. Descriptive statistics used to identify and compare the mean and
standard deviation of related questions was additionally used to demonstrate internal consistency.
The two questions included to intentionally address the same construct yielded very similar
responses throughout the survey.

<table>
<thead>
<tr>
<th>JCSEE Evaluation Standard</th>
<th>Question Numbers</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propriety Standard</td>
<td>1, 15, 14, 16</td>
<td>0.787 / Acceptable Range</td>
</tr>
<tr>
<td>Utility Standard</td>
<td>2, 6, 3, 4, 12, 13, 19, 20, 25, 26</td>
<td>0.751 / Acceptable Range</td>
</tr>
<tr>
<td>Feasibility Standard</td>
<td>8, 9, 10, 11</td>
<td>0.774 / Acceptable Range</td>
</tr>
<tr>
<td>Accuracy Standard</td>
<td>5, 7, 17, 18, 21, 22, 23, 24, 27, 28, 29, 30</td>
<td>0.868 / Good Range</td>
</tr>
</tbody>
</table>

The relative infancy of evaluation programs using student performance data may prompt
increased research in this area. The survey instrument constructed, piloted, used, and analyzed in
this study may be of value to researchers conducting similar research in the perceptions of
teachers to the use of specific student performance data in teacher evaluations.

Implications for Improving Teacher Evaluation and Instruction

The literature on teacher evaluation is full of examples of how current evaluation
programs are flawed. Teachers, administrators, and policymakers point out problems with
current evaluation programs that range from criticisms regarding the fidelity of the process to the
fact that evaluation results rarely effectively influence personnel or instructional outcomes
(Weisberg et al., 2009). Since the research clearly demonstrates that the quality of teaching matters, it is reasonable to presuppose that a quality teacher evaluation process also matters in order to know if the school system possesses high quality teachers (Stronge & Tucker, 2003). Teacher evaluations, therefore, need to acknowledge student achievement data. State policymakers and instructional leaders must similarly acknowledge the perceptions of teachers to the use of student performance data in their evaluations.

The fact that teacher perceptions in the study are not entirely favorable toward an evaluation program that uses student performance data certainly warrants further investigation. There is considerable evidence to document how an effective teacher is the single most important contributor to a student’s academic achievement gains (Goe et al., 2008; Wright, Horn, & Sanders, 1997). What does not appear as definitive to teachers in the study is the teachers’ confidence that student performance data accurately documents their effectiveness as a teacher. Teachers, for example, have regularly argued against the use of student performance data because it fails to recognize the inherent differences in every classroom and every school (Kelsey, 2009; Sand, 2005). In this study, teachers across a wide range of demographic factors report that the use of student performance data will not significantly improve the evaluation process. This perception changes significantly after teachers have experience in an evaluation program that uses student performance data.

The revelation that teachers with experience in an evaluation program that uses student performance data largely agreed with the premise that student performance data leads to more accurate and useful evaluations has at least three important implications for improving teacher practice. These implications include: (a) transforming evaluation programs into mechanisms for meaningful individual and school-wide professional growth, (b) utilizing student performance
data to drive responsible decision-making in schools to promote academic achievement for all students and to close achievement gaps, and (c) using evaluations to recognize exemplary teaching and make more informed personnel decisions and placements.

Teachers experienced with an evaluation program that uses student performance data were more optimistic toward the inclusion of student performance data in their evaluation. They pointed to the opportunities associated with this evaluation format to promote meaningful professional development. Teachers’ comments that a focus on student achievement data may actually facilitate personal growth and development supported other studies (Baker et al., 2010). In addition to the strong level of agreement noted by teachers to survey questions associated with the utility standards, teachers’ coded open-ended feedback specifically mentioned that student performance data would “inform professional development” and “enhance personal growth and reflection.” Other teachers noted that student performance data offered administrators and teachers a neutral and objective source of information that can launch constructive conversations between both parties. Coded responses that student performance data in their evaluation would “remove evaluator bias” and “increase collaboration” suggests that student achievement data in the evaluation instrument possesses the potential to drive meaningful dialogue between teacher and administrator. This supports research that this type of constructive dialogue has the opportunity to significantly assist the teacher’s professional growth when framed in the proper perspective (Castillo, 2005; Lyon, 2010).

Evaluation programs that include student performance data are better positioned to design and implement data-driven decision-making in schools that promote academic achievement and close achievement gaps. In the additional feedback section, teachers from across demographic categories cited numerous ways that evaluation programs that use student performance data
could promote academic achievement (see Appendix E). Additional feedback from teachers that were coded and grouped into categories such as “identifies student gaps,” “promotes data-driven-decision making,” and “guides lesson planning.” These comments reflect teachers’ attitudes that student performance data in evaluation programs can help identify and promote the use of instructional strategies that clearly increase student achievement. This can be of particular benefit in settings where achievement gaps between student groups are present. The coded feedback was supported by survey responses from teachers experienced with using an evaluation program that uses student performance data. Previous studies support the findings of this study. The use of student achievement data, for example, assists teachers and school leaders in the identification of student achievement gaps both at the classroom and district level and promotes the use of effective teaching strategies (Radmin, 2012; Sevillano, 2002). As student performance data is tied to teacher evaluations, teachers and administrators will be forced to examine and utilize the data to enhance teaching and learning.

A final implication for improving teaching through evaluation programs that use student performance data is through the program’s ability to recognize exemplary teaching and make more informed personnel decisions and placements. Teachers commented in the additional feedback section that the use of student performance data in the evaluation process would “identify good teachers” and “make teachers more accountable.” Teachers with at least one year of experience working with an evaluation program that uses student performance data also largely agreed to survey questions correlated with the propriety standard. This level of agreement suggests that these experienced teachers value this type of evaluation’s program to fairly differentiate between levels of performance. The teachers’ perceptions support the findings from a recent study that similarly criticized current evaluation systems as being
“disrespectful to teachers” and indifferent to instructional effectiveness (Weisberg, Sexton, Mulhern, & Keeling, 2009, p. 4). This same study noted that teachers have been routinely rated as satisfactory and above for decades. Truly effective teachers deserve to be distinguished from their colleagues and an evaluation program that uses student performance data can do this according to teachers who have participated in such a program. The instructional strategies employed by these outstanding educators can be more readily replicated in other classrooms.

**Implications from Teacher Homogeneity of Responses**

Results in this study demonstrated a high degree of homogeneity of perceptions among teachers across demographic factors. There is evidence that a teacher’s perceptions as to what constitutes effective teaching is influenced by personal experience (Murphy, Delli, & Edwards, 2004; Snider & Roehl, 2007). It appears that situational factors associated with personal experience play a minimal role in shaping perceptions regarding the use of student performance data in teacher evaluations. The only factor that did influence teacher perceptions was to what degree the teacher had experience with an evaluation program that uses student performance data. This was especially noticeable in the questions associated with the propriety and accuracy of evaluation standards. Since responses were not dependent on demographic factors, efforts to educate teachers about the use of student performance data in evaluations should focus on these common perceptions. It appears that teachers will demonstrate initial reluctance toward any new change to their evaluation program. Instructional leaders need to be prepared for this level of anxiety from all teachers. Instructional leaders, however, should also emphasize to teachers how the level of support toward an evaluation program that utilizes student performance data increases among teachers after they have one year of experience with such a program.
The homogeneity of teachers' responses should not prevent leaders from ignoring important differences in how to effectively deliver guidance and information regarding how student performance data will impact teachers. Although teachers from tested and non-tested courses responded similarly in this survey, leaders should consider differentiating their message to these two unique audiences since student performance data will presumably be captured differently. It is also important to consider differentiating the message to novice and veteran teachers. Again, this study noted very similar responses between teachers from varying levels of teaching experience on using student performance data in teacher evaluations. Teachers who have recently graduated from an education program, however, are more likely to have greater exposure to assessment and data courses. This relatively new concentration in undergraduate education preparation programs may help them better understand and appreciate the power of student performance data.

Concerns from Teachers

In this study, teachers expressed a number of concerns regarding the use of student performance data in their evaluations. The survey responses from teachers across all demographic levels indicate a general level of disagreement toward the use of student performance data. Their level of disagreement was reinforced by responses in the additional feedback section. The higher frequency of negative comments indicated that teachers opposed to the utilization of student performance data were more inclined to voice these feelings in more detail in this section of the survey. Limitations cited by teachers were coded into categories that included: fear of teaching to the test; student apathy toward tests (which would lead to inaccurate data); inaccurate assessments; and student performance data being the sole source of a teacher’s
evaluation (see Appendix E). It is important to note, however, that only two of the 34 teachers who have at least one year of experience with such a program cited any limitations.

It appears from this study that teachers without any experience or an informed background into the use of student performance data in teacher evaluations cite the same type of concerns reported in research that is now at least ten years old. Teachers complained then that a singular focus on a one-time assessment mitigates the other dimensions of a child’s acquisition and demonstration of learning and an overreliance on testing data compromised educational quality by leading teachers to teach to the test, focusing their classes on narrow test-taking strategies rather than on broader, conceptual material (Carpenter, 2001). Other criticisms of student performance data argued that this data did little to measure how much students actually were learning or how advanced their skills were” (Walker, 2000). Teachers in this survey without any experience teaching under an evaluation program that uses student performance data reiterated the same concerns cited in earlier research.

It is important to note that this decade-old research that may be influencing teacher perceptions is often based on criticisms of decade-old assessments. More recent research indicates that when tests were used properly, they served as a valuable and valid tool to measure student achievement (Knight, 2008). The increased use of assessments has led to more accurate and reliable assessments which may be one reason why teachers who have actually been evaluated using student performance data are more confident in the program’s ability to accurately evaluate their performance. The multiple-choice, closed-ended question formats found on most assessments have improved to test beyond the knowledge and comprehension levels of Bloom’s Taxonomy. The most cited concern by teachers was the fear that this would promote “teaching to the test.” Again, more recent studies discovered that component,
productive, and accountable teachers who generally excelled treated state tests as "nothing more than another useful guide and motivator, with no significant change" in the way they presented lessons to their students (Mathews, 2006, p.1). It is important for leaders to use this more updated research and findings from teachers who have used an evaluation program that uses student performance data in this study to alleviate the fears of teachers.
Conclusion

Teacher evaluation remains a significant element in ensuring effective teaching and learning is taking place in America’s schools. Instructional leaders are now recognizing the value in student performance data and the power inherent in this data to improve student learning outcomes (Sevillano, 2002). Student performance data is also a proven source of information for identifying potential gaps in instruction and academic achievement. The premise behind utilizing student performance data in teacher evaluations is the documented need for ensuring every child in every state receives a high-quality education. These “rigorous, transparent and fair evaluations systems for teachers” must take into account data on student growth (USDOE, 2010). The current study contributed to this task by identifying teachers’ perceptions to the use of student performance data in teacher evaluations.

The findings from the study indicate that teachers without previous experience with an evaluation program that includes student performance data are expressing initial levels of reluctance to embrace student performance data in their evaluations. Learning more about how teachers perceive the evaluation process is important since evaluations have not historically had the power to enhance teaching and learning. Since an evaluation is viewed by teachers as a significant part of his or her continued employment or teaching assignment, it is imperative for teachers to believe the evaluation is a reliable and valid indicator of their performance. Results from this study indicate that there is currently a common perception among teachers regarding the use of student performance data in teacher evaluations. Teacher perceptions were consistent among teachers across the demographic factors (a) years of experience, (b) union or non-union state; (c) tested or non-tested grade and course; and (d) elementary, middle, or high school.
There is an increasing literature base on teacher perceptions of teacher evaluations. Most of the research conducted on teachers’ perceptions sought to discover how teachers felt about their current evaluation program and whether it was a meaningful experience for the teacher (Breedlove, 2011; Clayton, 2008; Clemetsen, 2000; Davis, 2000; Doerr, 2012; King, 2003; Marks, 2005; Phillips, 2005; Sutton, 2008). These studies primarily explored how teachers perceived their evaluation process primarily through the use of qualitative studies that often utilized open-ended interviews and only marginally included some quantitative findings. Despite this abundance of research on teachers’ perceptions of their evaluation, these studies failed to specifically address how teachers perceive the use of student achievement data in their evaluation. A recent query (May 15, 2013) on the Education Research Complete database produced 285 matches for a combined search of the terms “teacher perceptions” and “teacher evaluations.” The results dropped significantly to 15 when a third term of “student performance” or “student achievement” was added to the search query. This study adds to the literature by identifying teacher perceptions to teacher evaluations that include student performance data.

**Delimitations and Limitations**

The study’s generalizability is affected by a number of factors. The study only included responses from K-12 public school teachers. Private school and charter school teachers were not part of the study’s sample. Since the percentage that student performance data factors into a teacher’s overall evaluation varies among states, teachers’ perceptions may differ. Teachers where student performance data accounts for 60% of the teacher’s evaluation may have responded differently from teachers in a state where student performance data only accounts for 20% of their evaluation. The study utilized a 30 question survey for data collection. The wording of the survey questions may also have affected teachers’ responses and perceptions.
The low response rate from the survey (n=166, 3.03%) also reduced the power of the findings. A larger sample size would have decreased the standard error of difference in the analyses (Kiess & Green, 2010).

**Recommendations for Future Research**

Additional research may add to the understanding of teachers’ perceptions of the use of student performance data in teacher evaluations. They following are recommended.

- In the current study, teachers were asked to provide their perceptions to the use of student performance data in teacher evaluations. For 80% of the respondents in this study, teachers had either no experience (n=90 or 54%) or were in the first year (n=42 or 26%) of such of such an evaluation program. It would be of value to replicate this study after teachers had more experience with this type of evaluation program. A study conducted five years after teachers had been exposed to an evaluation program that included student performance data might yield different results.

- It would be interesting to learn if teacher licensure type as an independent variable would affect the teachers’ perceptions to the use of student performance data in teacher evaluations. Since school districts across the nation are recruiting teachers from career-switcher programs to fill hard-to-staff teaching assignments, it would be interesting to learn if teachers with previous career experience in the business or military sector have similar perceptions about the use of student performance data in teacher evaluations.

- Further understanding of the current study could be achieved from focus group interviews that would follow survey administration. Focus groups could provide additional insight into how teachers perceive the use of student perception data in
teacher evaluations. If teachers are provided the opportunity to orally express their perceptions, there is a greater likelihood that errors associated with the wording or interpretation of survey questions of the current study are negated.

- Including the perceptions of parents, students, community leaders, and/or policymakers may prove to be a valuable source of data that would complement the current study. Since the accountability movement that pushes for a review of student performance is often led by policymakers, it would be interesting to note how their perceptions to the four evaluation standards compared to teacher perceptions. It would also be of particular interest to see if students find the use of their performance data to be of value in evaluating teachers. Teachers cited lack of student motivation and apathy toward tests to be a potential limitation to using student performance data. Student feedback might be able to support or contradict this concern cited from multiple teachers.
Washington, DC: Education Sector.


United States Department of Education. (n.d.). Measuring student growth for teachers in non-


Appendix A: Final Teacher Perception Survey

Teacher Perception Survey:
Teacher Perceptions to the Use of Student Performance Data in Teacher Evaluation

BACKGROUND: A number of states are using student performance data as a component in a public school teacher’s evaluation. This study is attempting to identify how teachers perceive the use of student performance data in their evaluations. For the purposes of this survey, student performance data is defined as student achievement progress, as determined by multiple measures of learning and achievement, including, when available and applicable, student-growth data from the state departments of education.

DIRECTIONS: Based on your knowledge and experience, please respond to the following questions regarding your perception on the use of student performance data in teacher evaluation. Please indicate the extent to which you agree with each statement along a scale of Strongly Agree (4), Agree (3), Disagree (2), to Strongly Disagree (1).
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I believe the use of student performance data in my evaluation will improve the overall evaluation process.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2. I believe the use of student performance data in my evaluation will make the evaluation process more informative to me as an educator.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3. I believe the use of student performance data in my evaluation will prompt me to focus more on student assessment data.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4. I believe the use of student performance data in my evaluation will direct my attention to potential achievement gaps for students in my classroom.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5. I believe the use of student performance data in my evaluation will more accurately evaluate my teaching.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6. I believe the use of student performance data will improve how I deliver instruction to my students.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7. I believe the use of student performance data in my evaluation will make my evaluation more objective.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8. I believe the use of student performance data as one performance standard in my evaluation is a responsible use of student assessment data.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9. I believe my evaluation should include data on how my students performed on appropriate and valid performance assessments.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10. I believe the student performance data in my evaluation will become the primary indicator of my effectiveness as a teacher.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11. I believe the use of student performance data will negate other performance standards and variables that impact teaching and learning in my classroom.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12. I believe the use of student performance data in my evaluation will increase the likelihood that I will devote more of my instructional time to “teaching to the test.”</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13. I believe the use of student performance data in teacher evaluations will limit my use of exploratory/enrichment teaching activities that are not directly connected with tested material.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>14. I believe the use of student performance data in teacher evaluations will help administrators accurately evaluate teaching performance.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>15. I believe the teacher evaluation process will be more meaningful to me with the use of student performance data in the evaluation.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>16. I believe the use of student performance data in teacher evaluations will help administrators identify low-performing/ineffective teachers.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>17. I believe the use of student performance data will lead to a more accurate evaluation of my teaching.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18. I believe the use of student performance data will more accurately document my strengths and weaknesses as a teacher.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19. I believe a teacher evaluation process that DOES NOT include student performance data will lead to the development of meaningful content or grade-level specific professional development for me as a teacher.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20. I believe a teacher evaluation process that DOES NOT include student performance data will provide my evaluator/administrator with sufficient information to suggest meaningful content-specific professional development activities for me.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>21. I believe a teacher evaluation process that DOES NOT include student performance data accurately will identify for my administrator specific content areas where I can improve as a teacher.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>22. I believe a teacher evaluation process that DOES NOT include student performance data will accurately identify for my administrator specific content areas where I excel as a teacher.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>23. I believe a teacher evaluation process that DOES NOT include student performance data will accurately evaluate my colleagues' overall performance and effectiveness as teachers.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>24. I believe an evaluation process that DOES NOT include student performance data will provide school administrators with sufficient information to make informed personnel decisions.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>25. I believe the use of student performance data in my evaluation will lead to the development of meaningful content or grade-level specific professional development for me as a teacher.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>26. I believe the use of student performance data in my evaluation will provide my evaluator/administrator with sufficient information to suggest meaningful content-specific professional development activities for me.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Please list three items that you favor with respect to the use of student performance data in teacher evaluations.

1. 
2. 
3. 

Please list three items that you fear with respect to the use of student performance data in teacher evaluations.

1. 
2. 
3.
Teacher Demographic Questions:

1) Please answer the following question with respect to your experience with an evaluation program that utilizes student performance or assessment data.
   A) I have not worked in or been evaluated in a school district that uses student performance or achievement data in a teacher's evaluation.
   B) This is my first academic year working in a school district that uses student performance or achievement data in a teacher's evaluation.
   C) I have more than one year experience of working in a school district that uses student performance or achievement data in a teacher's evaluation.

2) Are you currently working in a school district that operates under a collective bargaining agreement?
   A) Yes
   B) No

3) What grade level do you currently teach? (If you teach multiple grade levels, please select the grade level that best describes your teaching assignment.)
   A) Elementary School
   B) Middle School
   C) High School

4) Do you teach at least one tested grade or course? (A tested course is defined as a course or grade level where students participate in a standardized end-of-year assessment.)
   A) Yes
   B) No

5) Which of the following best describes your teaching experience?
   A) 0-5 Years of Teaching Experience
   B) 6-10 Years of Teaching Experience
   C) 11+ Years of Teaching Experience
Appendix B

Consent for Participation

Please read the following Consent Agreement before proceeding with the survey.

I agree to participate in a dissertation study investigating the perceptions of K-12 teachers on the use of student performance data in teacher evaluations. The purpose of this study is to determine how teachers perceive the use of student performance data in teacher evaluations with respect to the four evaluation standards of the Joint Committee Personnel Evaluation Standards. I understand that my selection to participate in this study is the result of a random selection process conducted by a third party vendor whose involvement in the study is limited exclusively to selecting and distributing information to potential participants. I understand that the researcher is conducting this study to fulfill requirements of a doctoral program in Education Policy, Planning, and Leadership at The College of William and Mary in Williamsburg, Virginia.

As a participant, I understand that my involvement in the study is limited solely to taking an online survey. I understand that the survey requires me to indicate my level of agreement with various statements about the use of student performance data in teacher evaluations. As a participant in the study I will provide relevant demographic information used in the study to answer research questions. I understand none of the information collected will be used to reveal my identity as a participant or to link my responses with my identity.

The survey is composed of 30 items and two open-ended items. This survey may take approximately 10-15 minutes to complete. I further understand that I may request a copy of the study’s results from the researcher by sending an email requesting results to pthopk@email.wm.edu.

I understand that there may be minimal psychological discomfort directly involved with this research. Further, I understand that I do not have to answer every questions asked of me, and I am free to withdraw my consent and discontinue participating in this study at any time simply by discontinuing the study. If I have any questions or problems that arise in connection with my participating in the study, I should contact Dr. James Stronge, the project director at 757-221-2339 or jhstro@wm.edu. If I have any ethical concerns with the conduct of this study, I should contact Dr. Michael Deschenes, the chair of the Protection of Human Subjects Committee at The College of William and Mary at 757-221-2778 or mrdesc@wm.edu.

By taking this survey, I verify that I am at least 18 years of age, that I have received copy of this consent form, and that I consent to participate in this study and the tasks outlined above.
Appendix C Letter to Participants

**Teacher Perception Survey:**

**Use of Student Performance Data in Teacher Evaluation**

Thank you for agreeing to complete the survey on teacher’s perceptions to the use of student performance data in teacher evaluations. The survey should not take more than 10-15 minutes of your time. Please take a moment to read the Consent of Participation below that describes the study and its ethical safeguards. Once you have reviewed this information, please click the “Yes, I would like to Participate” button to proceed to the next page to begin answering the survey questions. At the end of the survey, you will have an opportunity to enter in an email address to receive a final copy of the study.

BACKGROUND: A number of states are using student performance data as a component in a public school teacher’s evaluation. This study is attempting to identify how teachers perceive the use of student performance data in their evaluations.

For the purposes of this survey, student performance data is defined as student achievement progress, as determined by multiple measure of learning and achievement, including, when available and applicable, student-growth data from the state departments of education.

DIRECTIONS: Based on your knowledge and experience, please respond to the following questions regarding your perception to the use of student performance data in your teaching evaluation. Please indicate the extent to which you agree with each statement along a scale of Strongly Agree (4), Agree (3), Disagree (2), to Strongly Disagree (1).

Thank you for your participation.
Appendix D: Permission to Use and Modify Surveys

June 7, 2012

To: Paul T. Hopkins

RE: Request to Adapt Survey from CSE Report 334, *Effects of Standardized Testing on Teachers and Learning—Another Look* by Joan Herman and Shari Golan

Dear Paul:

Thank you for your email of June 7, 2012 requesting permission to adapt the survey in CSE Report 334, *Effects of Standardized Testing on Teachers and Learning—Another Look* by Joan Herman and Shari Golan.

Permission is hereby granted for you to adapt and use the above survey for your dissertation and any subsequent publication as necessary. Please use the following acknowledgement in your dissertation or other publication.

The survey used in this research was adapted from CSE Report 334, *Effects of Standardized Testing on Teachers and Learning—Another Look* by Joan Herman and Shari Golan. It was adapted with permission from The National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Copyright © 1991 and by The Regents of the University of California as supported under the Institute of Education Sciences (IES), U.S. Department of Education.

Please contact me if you have any further questions or if I may be of any additional help.

Sincerely,

Ronald Dietel, Ed.D.
Assistant Director for Research Use and Communications
CRESST/UCLA, 300 N. Charles Young Drive, #321
Los Angeles, CA 90095
310-794-9168
dietel@cse.ucla.edu
Dear Mr Hopkins

I do not have a problem. Unfortunately I did not keep the survey.

Best of luck.
Ansie Lessing

-----Original Message-----
From: Paul Hopkins [mailto:phopkins@advancepath.com]
Sent: 25 May 2012 22:03
To: Lessing, Ansie; de Witt, Maria
Subject: Permission to Adapt Survey

Good afternoon:

I came across your article, "The value of continuous professional development: teacher perceptions." I am wondering if I may be granted permission to adapt your survey instrument for the purposes of my dissertation work on teacher perceptions toward teacher evaluation. I will need to adapt the questions to relate more to teacher evaluation than professional development but I would like to follow your structure and wording if possible. Please let me know if you have any questions.

I will certainly credit you if I am granted permission and I decide to continue with my current plan for my dissertation.

Thank you for your consideration.

Paul T. Hopkins
Appendix E

Disaggregated Teacher Additional Feedback on Benefits of Student Performance Data in Teacher Evaluations

### Years of Experience

<table>
<thead>
<tr>
<th>0-5</th>
<th>6-10</th>
<th>11+</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Guides Lesson Planning (4)</td>
<td>• Guides Lesson Planning (2)</td>
<td>• Informs Professional Development (3)</td>
</tr>
<tr>
<td></td>
<td>• Data-Driven-Decision Making (2)</td>
<td>• Alignment to Common Core</td>
</tr>
<tr>
<td>• Identifies Student Gaps (3)</td>
<td>• Identifies Good Teachers</td>
<td>• Removes Evaluator Bias (2)</td>
</tr>
<tr>
<td>• Enhances Personal Growth and Reflection (2)</td>
<td>• Identifies Student Gaps (2)</td>
<td>• Increases Collaboration (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Makes Evaluation More Objective (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Identifies Student Gaps (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Makes Teachers Accountable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Teaches Test-Taking Skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Enhances Personal Growth and Reflection</td>
</tr>
</tbody>
</table>

### Working under a Collective Bargaining Agreement (CBA)

<table>
<thead>
<tr>
<th>Yes, Working under a CBA</th>
<th>No, Not Working under a CBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Guides Lesson Planning (3)</td>
<td>• Guides Lesson Planning (3)</td>
</tr>
<tr>
<td>• Identifies Student Gaps (5)</td>
<td>• Data-Driven-Decision Making (2)</td>
</tr>
<tr>
<td>• Enhances Personal Growth and Reflection</td>
<td>• Identifies Student Gaps (2)</td>
</tr>
<tr>
<td>• Identifies Good Teachers</td>
<td>• Enhances Personal Growth and Reflection (2)</td>
</tr>
<tr>
<td>• Informs Professional Development (2)</td>
<td>• Informs Professional Development</td>
</tr>
<tr>
<td>• Increases Collaboration</td>
<td>• Alignment to Common Core</td>
</tr>
<tr>
<td>• Makes Evaluation More Objective (2)</td>
<td>• Increases Collaboration</td>
</tr>
<tr>
<td>• Removes Evaluator Bias</td>
<td>• Makes Evaluation More Objective</td>
</tr>
<tr>
<td></td>
<td>• Removes Evaluator Bias</td>
</tr>
<tr>
<td></td>
<td>• Makes Teachers Accountable</td>
</tr>
<tr>
<td></td>
<td>• Teaches Test-Taking Skills</td>
</tr>
</tbody>
</table>

### Teaching a Tested Subject or Grade

<table>
<thead>
<tr>
<th>Yes, Teaching a Tested Subject or Grade</th>
<th>No, Not Teaching a Tested Subject or Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Guides Lesson Planning (5)</td>
<td>• Guides Lesson Planning (1)</td>
</tr>
<tr>
<td>• Identifies Student Gaps (6)</td>
<td>• Identifies Student Gaps (1)</td>
</tr>
<tr>
<td>• Enhances Personal Growth and Reflection</td>
<td>• Enhances Personal Growth and Reflection (2)</td>
</tr>
<tr>
<td>• Identifies Good Teachers</td>
<td>• Informs Professional Development</td>
</tr>
<tr>
<td>• Informs Professional Development (2)</td>
<td>• Increases Collaboration</td>
</tr>
<tr>
<td>• Increases Collaboration</td>
<td>• Makes Teachers Accountable</td>
</tr>
<tr>
<td>• Makes Evaluation More Objective (3)</td>
<td>• Teaches Test-Taking Skills</td>
</tr>
<tr>
<td>• Removes Evaluator Bias (2)</td>
<td></td>
</tr>
</tbody>
</table>

180
- Data-Driven-Decision Making (2)
- Alignment to Common Core

### Level of School

<table>
<thead>
<tr>
<th>ELEMENTARY</th>
<th>MIDDLE</th>
<th>HIGH</th>
</tr>
</thead>
</table>
| - Guides Lesson Planning (3)  
- Identifies Student Gaps (3)  
- Enhances Personal Growth and Reflection (2)  
- Identifies Good Teachers  
- Informs Professional Development  
- Increases Collaboration  
- Makes Evaluation More Objective (3)  
- Data-Driven-Decision Making (2)  
- Removes Evaluator Bias | - Guides Lesson Planning  
- Identifies Student Gaps  
- Informs Professional Development  
- Increases Collaboration  
- Makes Teachers Accountable | - Guides Lesson Planning (2)  
- Identifies Student Gaps (3)  
- Enhances Personal Growth and Reflection  
- Informs Professional Development  
- Teaches Test-Taking Skills  
- Alignment to Common Core  
- Removes Evaluator Bias |

### Participation in an Evaluation Program that Uses Student Performance Data (SPD)

<table>
<thead>
<tr>
<th>No Experience</th>
<th>First Year of Evaluation Program that Uses SPD</th>
<th>More than One Year of Experience with Evaluation Program that Uses SPD</th>
</tr>
</thead>
</table>
| - Guides Lesson Planning  
- Increases Collaboration  
- Makes Teachers Accountable | - Guides Lesson Planning (3)  
- Identifies Student Gaps (3)  
- Enhances Personal Growth and Reflection (2)  
- Informs Professional Development  
- Increases Collaboration (2)  
- Makes Evaluation More Objective (3)  
- Data-Driven-Decision Making (2) | - Guides Lesson Planning (2)  
- Identifies Student Gaps (4)  
- Enhances Personal Growth and Reflection  
- Informs Professional Development  
- Teaches Test-Taking Skills  
- Alignment to Common Core  
- Removes Evaluator Bias (2)  
- Identifies Good Teachers |

(#{}) – Number of times item appeared as an additional feedback response
Disaggregated Teacher Additional Feedback on Limitations of Student Performance Data in Teacher Evaluations

**Years of Experience**

<table>
<thead>
<tr>
<th></th>
<th>0-5</th>
<th>6-10</th>
<th>11 +</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>Becomes Sole Source of Evaluation (3)</td>
<td>Becomes Sole Source of Evaluation (5)</td>
<td>Becomes Sole Source of Evaluation (6)</td>
</tr>
<tr>
<td></td>
<td>Teaching to the Test (3)</td>
<td>Teaching to the Test (6)</td>
<td>Teaching to the Test (7)</td>
</tr>
<tr>
<td></td>
<td>Does not account for student ability groupings (2)</td>
<td>Does not account for student ability groupings (1)</td>
<td>Does not account for student ability groupings (2)</td>
</tr>
<tr>
<td></td>
<td>Does not recognize uniqueness of school population (4)</td>
<td>Does not recognize uniqueness of school population (2)</td>
<td>Does not account for student ability groupings (2)</td>
</tr>
<tr>
<td></td>
<td>Teaching Students with Disabilities or English Language Learners</td>
<td>Teaching Students with Disabilities or English Language Learners</td>
<td>Student Apathy Toward Test (7)</td>
</tr>
<tr>
<td></td>
<td>Student Apathy Toward Test</td>
<td>Student Apathy Toward Test (2)</td>
<td>Student Apathy Toward Test (7)</td>
</tr>
<tr>
<td></td>
<td>Inaccurate Assessments</td>
<td>Misuse of Data by Administrator (2)</td>
<td>Misuse of Data by Administrator (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inaccurate Assessments (2)</td>
<td>Inaccurate Assessments (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Working under a Collective Bargaining Agreement (CBA)**

<table>
<thead>
<tr>
<th>Yes, Working under a CBA</th>
<th>No, Not Working under a CBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becomes Sole Source of Evaluation (6)</td>
<td>Becomes Sole Source of Evaluation (8)</td>
</tr>
<tr>
<td>Teaching to the Test (7)</td>
<td>Teaching to the Test (9)</td>
</tr>
<tr>
<td>Does not account for student ability groupings (3)</td>
<td>Misuse of Data by Administrator (3)</td>
</tr>
<tr>
<td>Does not recognize uniqueness of school population (4)</td>
<td>Student Apathy Toward Test (6)</td>
</tr>
<tr>
<td>Teaching Students with Disabilities or English Language Learners</td>
<td>Does not account for student ability groupings (2)</td>
</tr>
<tr>
<td>Student Apathy Toward Test (4)</td>
<td>Does not recognize uniqueness of school population (6)</td>
</tr>
<tr>
<td>Misuse of Data by Administrator (4)</td>
<td>Inaccurate Assessments (2)</td>
</tr>
<tr>
<td>Inaccurate Assessments (3)</td>
<td>Teaching Students with Disabilities or English Language Learners (2)</td>
</tr>
</tbody>
</table>
### Teaching a Tested Subject or Grade

<table>
<thead>
<tr>
<th>Yes, Teaching a Tested Subject or Grade</th>
<th>No, Not Teaching a Tested Subject or Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Becomes Sole Source of Evaluation (11)</td>
<td>• Becomes Sole Source of Evaluation (3)</td>
</tr>
<tr>
<td>• Teaching to the Test (12)</td>
<td>• Teaching to the Test (4)</td>
</tr>
<tr>
<td>• Does not account for student ability groupings (4)</td>
<td>• Misuse of Data by Administrator (5)</td>
</tr>
<tr>
<td>• Does not recognize uniqueness of school population (7)</td>
<td>• Student Apathy Toward Test (4)</td>
</tr>
<tr>
<td>• Teaching Students with Disabilities or English Language Learners (3)</td>
<td>• Does not account for student ability groupings</td>
</tr>
<tr>
<td>• Student Apathy Toward Test (6)</td>
<td>• Does not recognize uniqueness of school population (3)</td>
</tr>
<tr>
<td>• Misuse of Data by Administrator (2)</td>
<td>• Inaccurate Assessments (3)</td>
</tr>
<tr>
<td>• Inaccurate Assessments (2)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of School</th>
<th>ELEMENTARY</th>
<th>MIDDLE</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Becomes Sole Source of Evaluation (6)</td>
<td>• Becomes Sole Source of Evaluation (3)</td>
<td>• Becomes Sole Source of Evaluation (5)</td>
</tr>
<tr>
<td></td>
<td>• Teaching to the Test (5)</td>
<td>• Teaching to the Test (4)</td>
<td>• Teaching to the Test (7)</td>
</tr>
<tr>
<td></td>
<td>• Does not account for student ability groupings (4)</td>
<td>• Does not account for student ability groupings (2)</td>
<td>• Misuse of Data by Administrator (7)</td>
</tr>
<tr>
<td></td>
<td>• Does not recognize uniqueness of school population (4)</td>
<td>• Does not recognize uniqueness of school population (2)</td>
<td>• Student Apathy Toward Test (7)</td>
</tr>
<tr>
<td></td>
<td>• Teaching Students with Disabilities or English Language Learners</td>
<td>• Teaching Students with Disabilities or English Language Learners</td>
<td>• Does not account for student ability groupings</td>
</tr>
<tr>
<td></td>
<td>• Does not recognize uniqueness of school population (2)</td>
<td>• Student Apathy Toward Test (3)</td>
<td>• Does not recognize uniqueness of school population</td>
</tr>
<tr>
<td></td>
<td>• Misuse of Data by Administrator (2)</td>
<td>• Misuse of Data by Administrator (2)</td>
<td>• Inaccurate Assessments (4)</td>
</tr>
<tr>
<td></td>
<td>• Inaccurate Assessments</td>
<td>• Inaccurate Assessments</td>
<td>• Teaching Students with Disabilities or English Language Learners</td>
</tr>
</tbody>
</table>
### Participation in an Evaluation Program that Uses Student Performance Data (SPD)

#### No Experience with Evaluation Program that Uses SPD
- Becomes Sole Source of Evaluation (8)
- Teaching to the Test (9)
- Does not account for student ability groupings (4)
- Does not recognize uniqueness of school population (3)
- Teaching Students with Disabilities or English Language Learners (2)
- Inaccurate Assessments (2)
- Misuse of Data by Administrator (5)
- Student Apathy Toward Test (2)

#### First Year of Evaluation Program that Uses SPD
- Becomes Sole Source of Evaluation (4)
- Teaching to the Test (5)
- Does not account for student ability groupings (2)
- Teaching Students with Disabilities or English Language Learners (2)
- Does not recognize uniqueness of school population (5)
- Student Apathy Toward Test (5)
- Misuse of Data by Administrator (5)
- Inaccurate Assessments (2)

#### More than One Year of Experience with Evaluation Program that Uses SPD
- Becomes Sole Source of Evaluation (2)
- Teaching to the Test (5)
- Misuse of Data by Administrator (5)
- Student Apathy Toward Test (3)
- Does not account for student ability groupings (3)
- Does not recognize uniqueness of school population (2)
- Inaccurate Assessments (2)

(#) – Number of times item appeared as an additional feedback response
Vita

Paul Thomas Hopkins

Birthdate: March 16, 1973
Birthplace: Springfield, Missouri

Education:

2013 The College of William and Mary
Williamsburg, Virginia
Doctor of Education, EPPL

2007 The College of William and Mary
Williamsburg, Virginia
Master of Education, K-12 Administration

1995 Vanderbilt University
Nashville, Tennessee
Bachelor of Science, History & Secondary Education

Publications and Presentations:


Academic and Professional Education Positions:

Director of Curriculum & Vice President of Academy Operations (2012-present). AdvancePath Academics, Williamsburg, Virginia.

Director of Exceptional Education (2010-12). Henrico County Public Schools, Henrico, Virginia.

Principal (2008-10). Grafton High School, Yorktown, Virginia.

Assistant Principal (2006-08). York High School, Yorktown Virginia.


History Department Chair, Interim Dean of Middle School, and Teacher (2000-2005). Hampton Roads Academy, Newport News, Virginia.
