2016

What Are The Professional Development Needs Of Principals As They Implement The New Educator Effectiveness System?

Susan L. Salvesen
Lehigh University

WHAT ARE THE PROFESSIONAL DEVELOPMENT NEEDS OF PRINCIPALS AS THEY IMPLEMENT THE NEW EDUCATOR EFFECTIVENESS SYSTEM?

By

Susan L. Salvesen

Presented to the Graduate and Research Committee

Lehigh University

Candidacy for the Degree of

Doctor of Education

Department of Educational Leadership

Lehigh University

September 2015

Certificate of Approval

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Education.

_____
Date

_____
Dissertation Chair
Dr. George P. White

_____
Accepted Date

Committee Members:

_____
Dr. Floyd D. Beachum

_____
Dr. Louise E. Donohue

_____
Dr. Robert Hassler

Acknowledgements

My deepest appreciation goes to my husband, Erling, whose unending love and support enabled me to reach the pinnacle of this journey. Without his patience and gentle prodding I would never have made it this far.

Special thanks to my children, Skip, Jedd, and Blake for their encouragement. The 3:00 AM tech assistance, the goodie box, and those power visits with Crystal, Eli, and Elliana Lynn, gave me that extra boost to keep on going. Thanks, too, to my sister, Aundrea and her mentor, Ron, who would not let me give up and to my brother, Jody, for his unconditional love.

I am forever grateful to my parents, Bettie and Joe Brown, for teaching me about grit, perseverance, and the value of hard work. And along with my parents, my Central Bucks coaches, Peg Solitario and Bill Walter, believed in me and encouraged me to reach beyond my comfort zone.

Much gratitude goes to my committee chair, Dr. George White, for his guidance, and many edits in this rigorous process. His creativity and depth of knowledge gave me much to think about to make my study as meaningful as possible. Thanks, too, to my committee, Dr. Louise Donohue, Dr. Floyd Beachum, and Dr. Bob Hassler, for their time, expertise, and the value they put on my study.

Finally, I would like to give a hug to my devoted Piggy who sat next to me in the cold study and walked with me to clear my head through this whole ten-year process. We made it!

Table of Contents

Appendices

List of Tables

List of Figures

Abstract

With the passing of Act 82, the state of Pennsylvania has provided school districts with Danielson's Framework as a tool for principals to evaluate teachers. The purpose of this study was to determine the perceived professional development needs of Pennsylvania principals as they implemented the new educator effectiveness system. Three hundred principals from across the state participated in the study.

The findings of this study suggest that principals who participated in this study were confident in assessing the elements in Domains 2 and 3 of the Framework. Principals have the most confidence in evaluating Component 2d: Managing Student Behavior and 2a: Creating an Environment of Respect and Rapport. The component principals have the least confidence evaluating is 3d: Using Assessment in Instruction. When the elements were examined in unconventional, but logical groupings, it was found that assessment in instruction continued to be the grouping that principals had less confidence in evaluating. Elements associated with student ownership of the learning were the elements that principals had the least confidence in evaluating. This study extended the literature on teacher evaluation by recognizing that principals are confident in evaluating teachers using the Framework. Evaluating teachers based on the actions of their students as indicated through classroom observations may be an area to examine in more depth.

This study also extended the literature base by identifying the preferred professional development formats in which principals would be most willing to participate. It was found that district and I.U. sponsored workshops were the preferred format for professional development. University course work, whether on campus,

online, or a hybrid of the two, was the least favorable means of professional development for principals.

The findings suggest that those providing professional development for principals would do well to examine how principals are able to collect evidence to support the students' learning and participation in the learning process. Also, professional developers, including school districts, intermediate units, and universities, should offer professional development that is relevant to the demographic population through workshops, mentoring/coaching sessions, or small study groups.

CHAPTER I

Introduction

The purpose of the reauthorization of the 1965 Elementary and Secondary Education Act (ESEA) was "to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind" (U. S. Department of Education, 2002, sec. 1). This nationwide call for accountability, renamed No Child Left Behind (NCLB), was a continuation of the federal government's objective to improve our nation's educational system.

The federal government has been a dominant influence on public education (Kessinger, 2011) since the National Commission on Excellence in Education published its landmark publication, *A Nation at Risk: The Imperative for Educational Reform* (The National Commission on Excellence in Education, 1983). This publication asserted that a rising tide of mediocrity was eroding our educational system and that the federal government has the responsibility to supplement the state and local systems to reach higher educational goals. The reauthorization required states and local governments to develop policies and programs that support the requirements of NCLB. States established standards and created a testing system that would assess students in math and reading in grades three to eight and high school. All students were expected to meet or exceed state standards by 2014. As a result, NCLB has become the audit function of American schooling (O'Brien & Roberson, 2012). According to the legislators who wrote the law, test-based accountability, or the use of standardized testing to measure the success of student learning, would be evidence of greater accountability for teacher instruction.

Touted as a "silver bullet" approach (Booher-Jennings, 2006), this legislation, designed to assess students, was intended to grab the attention of educators and provide them with objective information that they could use to make instructional improvements (Diamond & Spillane, 2003). The goals of high-stakes testing were to improve student learning in essential academic content (reading and math), and to guarantee all students, not just equal access, but an equal education (McGee, 2005; Porter, 2000). In addition, principals and teachers should feel personal and collective responsibility for student achievement (Elmore, 2000; Vasquez-Heilig, Young, & Williams, 2011).

Role of the Principal

Studies show that test-based accountability has redirected the work of the principal and teacher by redirecting time, materials and personnel to emphasize worthwhile academic content (Rutledge, Harris, & Ingle, 2010; Anagnostopoulos & Rutledge, 2007; Ladd & Zelli, 2002). Teachers and students are working harder and teachers are working more efficiently and effectively (Porter, 2000). There are also some unintended and negative consequences. Some studies show that the sanctions and external threats associated with NCLB foster an environment of fear and embarrassment (Diamond & Spillane, 2004; Firestone, 2009; Vasquez-Heilig et al., 2011; Mintrop & Sunderman, 2009; Rice & Malen, 2003). The stringent test security measures, need for bathroom patrols, and scrutinization of erasure smudges are only some of the policing policies in place that cause educators to feel like offenders. Principals and teachers increasingly search for loopholes, such as: teaching to the bubble (those students who have the potential to pass the test), classifying more students for special education, or retaining students if they are not ready to take the test. In this environment, students are

viewed as liabilities, rather than opportunities (Booher-Jennings, 2005; Vasquez–Heilig, et al., 2011). The removal of principals and teachers for the lack of student achievement is a dramatic negative consequence.

The roles and responsibilities of the principal have taken a dramatic shift in this new culture of accountability. When examining daily logs of what principals actually do with their time, the research shows that management, personnel issues, and student affairs received a larger proportion of principals' attention than would be expected (Camburn, Spillane, & Sebastian, 2010). Principals can no longer be exclusively supervisors who are simply managing a building, budgeting, disciplining, or dealing with disruptions. The principal has become the pedagogical leader of the school (Elmore, 2000; Levine, 2005). Ravitch (2011) found that when low performing schools improved, much of the credit could be attributed to the work of the principal, along with a dedicated staff. Moreover, Leithwood (2008) asserted there was no documented case of a school successfully turning around its academic achievement in the absence of talented leadership. The spotlight of reform is clearly on school leadership. When using high-stakes testing as the criterion for student achievement, accountability for successful results falls to the principal (Levine, 2005). So, too, does the overall climate of the educational community.

In 2010, another level of accountability was thrust upon the principal. The U.S. Department of Education provided $4.35 billion for the Race to the Top Fund, which was targeted to improve student achievement. With this financial bonus, states were required to:

- develop a system for tracking student growth,

- design and implement an evaluation system for teachers and principals that would differentiate performance, taking student growth into account,
- conduct annual evaluations of teachers and principals,
- use the results of these evaluations to inform decisions, and
- provide relevant staff development for teachers and principals (U.S. Department of Education, 2009).

While many states already had a system in place for tracking student growth through standardized testing, the focus was now on designing and implementing an evaluation system that would include multiple measures to evaluate teacher effectiveness including student achievement and growth.

Traditional Evaluation System

The traditional measure of a teacher's effectiveness was a one-way, top down evaluation, which designated a teacher either satisfactory or unsatisfactory. Toch and Rothman (2008) describe the current teacher evaluation system as superficial and capricious; one that does not address quality of instruction, much less student learning. By using a binary system of satisfactory and unsatisfactory, teachers become part of the Widget Effect, one size fits all (Weisberg, Sexton, Mulhern, & Keeling, 2009). Led by Michele Rhee, The New Teacher Project examined twelve diverse school districts in four states and found that using the binary system creates the fallacy that all teachers are interchangeable. The results of the study showed that good teaching is ignored and unrecognized, bad teachers languish, and moderate teachers do not receive the staff development they need to improve. It was found that rather than a system that improves and sustains the quality of the teacher workforce, two factors that are vital priorities in

raising student achievement (Looney, 2011), the traditional system was used exclusively for decisions related to teacher remediation and dismissal (Weisberg, et al., 2009). It was a system that failed to assess the variations in effectiveness between teachers and also failed to diagnose specific staff development needs for teachers (Marshall, 2005).

Rationale for Change

Danielson (2012) provided four fundamental elements for a successful evaluation system designed to improve teacher quality. First, she claimed, it is essential to define what effective teaching looks like. Danielson's Framework is aligned with the core standards from the Interstate New Teacher Assessment and Support Consortium (INTASC) developed by The Council of Chief State School Officers (2011). Second, she claimed these standards must be shared with stakeholders so that those working with the standards clearly understand the expectations. She also claimed, this would help teachers to focus their reflections. According to Danielson, greater specificity of language would make it easier to pinpoint strengths and areas in need of improvement. Moreover, she stated, clear standards would provide a common language for collegial discussions. Third, she asserted a critical component of any evaluation system is in the skill of the evaluator to objectively analyze a teacher's performance, to interpret the performance in relation to the standard, and to involve the teacher in a reflective and productive conversation for effective solutions. Other scholars claimed that without a knowledgeable background, the evaluator would not be able to gain the trust to be able to deliver critical feedback (Donaldson & Donaldson, 2012). Finally, Danielson claimed an effective evaluation plan must be differentiated to meet the needs of each teacher.

By contrast to the traditional method of evaluation, a differentiated model of evaluation would be planned and systematic. The evaluation of a novice teacher may look different than that of an experienced teacher. The teacher evaluation system helps tease out the variation of skills of teachers. Used as a formative assessment, teacher evaluation can be used to align professional development opportunities to needs, while encouraging a professional community of learners (Looney, 2011; Hazi & Rucinski, 2009; PSEA, 2012; Goldrick, 2002). Formative evaluation can provide a personal growth plan for teachers (Donaldson & Donaldson, 2012). Donaldson and Donaldson (2012) warned that educators need to protect these opportunities to learn and grow. Furthermore, they claimed the success of a teacher evaluation system would depend on the allotment of time for study, the provision of resources, and trusting relationships.

In a recent study, Marzano (2012) surveyed 3,000 teachers as to the purpose of teacher evaluation. Seventy-six percent of surveyed teachers felt that the purpose of teacher evaluation was to measure and develop a teacher's effectiveness, but that the development of the teacher was most important.

According to the Gates Foundation, the goal of a new evaluation system was to build a fair and reliable program for teacher observations that would give feedback to the observer and the observed, provide an opportunity for professional development, and encourage continuous improvement. (Bill and Melinda Gates Foundation, 2010). Furthermore, it would provide a knowledge base for practitioners who are trying to strengthen the teaching profession and would be used as a tool for instructional improvement (Darling-Hammond, 2012).

While improving teaching quality was the initial goal of a new teacher evaluation plan, ultimately the objective will be to improve student learning. To quantify this, multiple measures have been used to evaluate teachers including student performance on standardized tests, professional portfolios, lesson plans, evaluation of student work, written reflections, video analysis, peer reviews, evidence of work with parents and peers, and evaluations by multiple evaluators over multiple occasions (Toch & Rothman, 2008). How evaluators define these measures will determine whether they are used to benefit or control teachers (Hazi & Racinski, 2009).

An independent report from the RAND Corporation surveyed 4,444 teachers (with a response rate of 81%) and 1,193 leaders (with a response rate of 76%) regarding the implementation of the Measures of Effective Teaching (MET) reform study in several school districts across the country (Stecher, Garet, Holtzman, & Hamilton, 2012). Following a full year of implementation of the reform, surveys showed that teachers believed that the new evaluation system had worthwhile goals of improving instruction, identifying areas for staff development, and determining the need for extra support. Eighty-eight percent of the teachers reported they were aware of the instructional expectations and seventy percent believed that the observers were well-qualified and provided useful feedback.

The strengths of the system have been the specificity of the assessment tool and the common language that have allowed for more meaningful conversations between school leaders and teachers, as well as among teachers. Teachers reported that they felt collegiality was enhanced rather than becoming more competitive. Ninety-five percent of the school leaders reported that they believed the system was fair and would

benefit students in the long run. Principals claimed they understood the connection between the teacher evaluation tool and professional development, and teachers saw the principal as the instructional leader. Unfortunately, it was reported that the professional development was not individually focused at this time. Stecher, et al., (2012) questioned what teachers' perceptions would be if their scores on the evaluation were to determine their placement or impact their compensation.

In anticipation of the requirements of Race to the Top, Pennsylvania Governor Corbett enlisted Team PA to conduct a pilot study during the 2010-2011 school year focused on developing a set of performance measures for improving the use of classroom observations and student data on the evaluation of teachers and principals (Lane & Horner, 2011). Team PA, funded in part by a Bill and Melinda Gates Foundation Grant, is a partnership of business leaders and senior government officials whose mission is to support Pennsylvania in being a national leader in education and economic development (Team PA Foundation, 2011a). The philosophy of Team PA according to CEO Matt Zieger is that by creating a quality evaluation system, just as in business, educators will be able to make informed decisions on "district human capital," (Team PA Foundation, 2011b, para. 11).

Similar to the results of the MET program, teachers and principals participating in this pilot study agreed that this new evaluation system supported quality instruction. However, principals and teachers felt that the training did *not* adequately prepare them for the implementation of the new system (Lane & Horner, 2011).

Measuring Educator Effectiveness

This pilot project paved the way as the State moved to incorporate a more effective teacher evaluation system. The State's response to the requirements for Race to the Top funds was Act 82 (Pennsylvania General Assembly, 2012). Signed into law in June of 2012 by Governor Corbett, Act 82 amended Public School Code of 1949. This legislation requires the Department of Education to develop a teacher and principal evaluation tool and requires multiple measures of educator effectiveness. Beginning with the 2013 school year, teachers' evaluations have been based on 85% observation and practice measured on Danielson's framework. Teachers have been receiving a score from 0-3 on each of the four domains of the framework. The remaining fifteen percent of the teachers' evaluation has been based on building level achievement data or PA School Performance Profile score. This will include PSSA or Keystone achievement, student performance on state value-added assessment calculations, graduation and/or promotion rate, attendance rate, AP course participation and student test data on the Scholastic Aptitude Test (SAT) and Preliminary Scholastic Aptitude Test (PSAT). In future years beginning in 2015, teacher ratings will be determined through teacher observation based on Charlotte Danielson's work, the School Performance Profile, a combination of scores based on whole school achievement, Teacher Specific Data consisting of longitudinal data collected over the next three years, and Elective Data based on district designed measures. Elective Data or Student Learning Objectives (SLOs) include district designed measures and examinations, nationally recognized standardized tests, industry certification examinations, student projects and/or portfolios pursuant to local requirements. (PDE, 2013a)

Unfortunately, there is limited empirical research to guide policymakers on how to combine these multiple indicators to achieve specific goals, such as: retaining the best teachers, planning for professional development that will enhance the performance of teachers, and supporting teacher evaluations (Milhaly, McCaffrey, Staiger, & Lockwood, 2013). The research done through the MET Project looked at the dimensions of teaching that were valued by experts and how these dimensions were measured. They also determined the optimal statistical weighting for predicting teacher performance and how this information is gathered from the data (Milhaly, et al., 2013). It remains to be seen how the composite estimator of teacher effectiveness correlates to student achievement.

Nevertheless, all classroom teachers will have a standard summative evaluation form from PDE that includes the four domains from the *Framework for Teaching*:

1. *Planning and preparation*: including selecting standards-based lesson goals and designing effective instruction and assessment;
2. *Classroom environment*: including establishing a culture for learning and appropriate classroom management techniques that maximize instructional time;
3. *Instruction*: including the use of research-based strategies which engage students in meaningful learning and utilize assessment results to make decisions about student needs;
4. *Professional Responsibilities*: including using systems for managing student data and communicating with families of students.

There are twenty-two components clustered into the Framework with each defining a distinct aspect of the domain. Each component has two to five elements that

further describe the component. While the components and elements are specific to a domain, they are related to each other as they affect different aspects of the complex activity of teaching (Danielson, 2007).

The PDE has developed rating scales and overall score ranges for the four performance levels (Pennsylvania General Assembly, 2012). Professional employees are required to be rated at least once every year, while temporary professional employees are to be rated at least twice annually. The summative evaluations will be reported as Distinguished, Proficient, Needs Improvement, or Failing. Only Failing is considered unsatisfactory. A second Needs Improvement rating within ten years of the first by the same employer would be considered an unsatisfactory rating. An Unsatisfactory and Needs Improvement rating would require the principal to develop a performance improvement plan for the teacher. Following four months on the plan, the teacher would be eligible for another evaluation. Two unsatisfactory ratings may lead to dismissal. Before an employee can be dismissed, the principal must provide a description of deficiencies in practice based on classroom observations and supported by detailed anecdotal records supporting a failing rating. At this time, there is no language in Act 82 requiring or prohibiting differentiated salary based on teacher performance ratings.

Furthermore, education specialists and non-teaching professionals will be rated under the Act 82 requirements beginning with the 2014-2015 school year. Professionals under these categories include: school counselor, dental hygienist, home and school visitor, instructional technology specialist, school nurse, and school psychologist. Eighty percent of these employee's ratings will be based on planning and preparation, educational environment, delivery of services, and professional development. School

performance of all students in the school building where the specialist and non-teaching professional is employed will comprise 20% of the overall rating. PDE is in the process of developing the criteria for this score (PDE, 2013b).

Rationale for the Study

With the passing of Act 82, the State has provided an evaluation tool that requires principals to do more than just produce a summative evaluation of a teacher as satisfactory or unsatisfactory. Formatively evaluating teachers will be much more challenging than providing a summative evaluation where the objective is minimum competencies (Brandt, 1987). Previously, PDE 428 (Appendix A) has been used to evaluate teachers. With the new Educator Effectiveness Program, the principal will be called upon to evaluate teachers using PA 82-1 (Appendix B). Each domain of the evaluation framework will be assessed to provide a numerical rating of 0 (Failing), 1 (Needs Improvement), 2 (Proficient) or 3 (Distinguished). These scores will be calculated along with the Building Level Rating, Teacher Specific Rating and Elective Rating to provide an overall score for each teacher. Ultimately, principals have been left to their own devices to determine how to implement the supervision of teachers in order to meet the requirements of the evaluation system (Spillane & Kenney, 2012). It is easy to delineate between the competent teacher and the incompetent teacher (Jacob & Lefgren, 2008; Brandt, 1987), but now principals will need to explain the requirements to differentiate between a proficient teacher and a distinguished teacher (Goe, Bell & Little, 2008). This brings a new challenge to principals.

The Framework will give principals and teachers a common language that will foster substantive dialogue as well as establish a shared understanding of effective

instruction. Due to the high inferential nature of teacher observation, more than other forms of evaluation, observation lends itself to evaluator bias that can threaten the validity and reliability of the protocol (Goe, 2009). Principals will need to be knowledgeable of the protocol and instructional methods in order to question a teacher's instruction and then be empathetic coaches delivering uncomfortable information (Donaldson & Donaldson, 2012). Ravitch (2011) suggested that principals who are evaluating teachers need to be master teachers themselves. Proper training for principals is essential not only in using the tool, but also in scoring videos of instruction so that there will be inter-rater reliability and all teachers will be getting the same score no matter who the evaluator is that conducts the observation (Goe, 2009; Bill and Melinda Gates Foundation, 2013).

Not only will the principal be required to determine the current level of performance of the teacher, he/she will need to pinpoint areas for growth and develop specific plans for improvement (Danielson, 2012; Brandt, 1987; Goe, 2009). In 2011 – 2012, billions of dollars were given to states under Title II of the Elementary and Secondary Education Act (ESEA), of which 44% of the money was allocated for professional development (U.S. Department of Education, 2012). Experts claim it is critical that the staff development offered to our teachers and principals is provided in meaningful ways that produce results (Jaquith, Mindich, Wei, & Darling-Hammond, 2011).

Principals need to make informed decisions regarding teacher recruitment, evaluation, professional development, placement, tenure, compensation, and retention. As the stakes get higher and accountability looms heavily on principals, it is increasingly

more important that principals be trained to make professional and legally defensible judgments as they evaluate teachers (Danielson, 2012). PSEA (2012) supports staff development for a teacher that is directly related to the teacher's job, driven by clear goals, is based on appropriate data, and is provided with input from the teacher. This custom designed staff development is a powerful way to improve teacher effectiveness. It is the role of the principal to differentiate the staff development of teachers to meet their individual needs.

In conclusion, this research will help to pinpoint the specific areas of growth needed for principals to refine their instructional leadership. Principals, who can guide their teachers in improving their instruction, will inevitably increase student learning. This research will also add to the existing research on professional development formats as well as personal and organizational characteristics of principals.

Purpose of the Study

The primary purpose of this study is to determine the perceived professional development needs of Pennsylvania principals as they implement the educator effectiveness system. In doing so, this study will examine the self-perceived readiness of principals to evaluate the components of the Danielson Rubric in the Domains 2 (The Classroom Environment) and 3 (Instruction). In addition, the study will determine whether the perceived professional development needs of principals are significantly related to personal demographics and organizational characteristics of the principals. These will include: the number of years of experience a principal is in this role, the number of teachers evaluated annually, the grade configuration of the school (elementary, middle, high school), the percentage of poverty level students identified by the

percentage of students receiving free and reduced lunch, the School Performance Profile percentage, and the model for teacher evaluation that the principal is using.

This study will also examine the formats principals consider important in determining their participation in professional development activities. The study will examine the relationship between the delivery of services, such as workshops, online courses, university classes, study groups, or conferences and principals' personal demographics and organizational characteristics.

Research Questions

The following research questions will guide this research study:

1. What is the confidence level of principals to evaluate teachers on the components of Domains (2) The Classroom Environment and (3) Instruction of the Danielson Framework?
2. What formats do principals prefer for participation in professional development activities related to implementing the educator effectiveness system?
3. What is the relationship between principals' personal demographics and their confidence regarding teacher evaluation?
    a. What is the relationship between principals' years of experience as a principal and their confidence regarding teacher evaluation?
    b. What is the relationship between the average number of teachers evaluated by the principals within a year and principals' confidence regarding teacher evaluation?
4. What is the relationship between organizational characteristics and principals' confidence regarding teacher evaluation?

a. What is the relationship between elementary, middle, and high school principals and their confidence regarding teacher evaluation?
b. What is the relationship between the percentage of free and reduced lunch students in a principal's school and the principal's confidence regarding teacher evaluation?
c. What is the relationship between principals' School Performance Profile percentage and their confidence regarding teacher evaluation?

5. What is the relationship between principals' personal and organizational characteristics and their preference regarding professional development formats related to implementing the educator effectiveness system?
    a. What is the relationship between principals' years of experience as a principal and their preferences regarding professional development formats?
    b. What is the relationship between the average number of teachers evaluated by principals within a year and their preferences regarding professional development formats?
    c. What is the relationship between elementary, middle, and high school principals and their preferences for professional development formats?
    d. What is the relationship between the percentage of free and reduced lunch students in principals' schools and principals' preferences for professional development?
    e. What is the relationship between principals' School Performance Profile percentages and their preferences for professional development?

Definition of Terms

Classroom Teacher – A professional or temporary professional employee who provides direct instruction to students related to a specific subject or grade level. (PDE, 2013a)

Formative Evaluation – A teacher performance appraisal designed to provide assistance to the teacher for the purpose of improving instruction.

Intermediate Unit - A regional educational agency that provides educational support to the students, parents, educators, and school administrators in a community.

Non-Teaching Professional Employee – A person who is an education specialist, professional or temporary employee who provides services other than classroom instruction. (PDE, 2013a)

Performance Improvement Plan – A plan designed by the principal with input from the employee that may include mentoring, coaching, recommendations for professional development, and intensive supervision based on the rating tool for ratings of failing and needs improvement. (PDE, 2013a)

Principal – An individual who is certified as a building principal, an assistant principal, a vice principal or a director of vocational education. (PDE, 2013a)

School Performance Profile – Developed by PDE to provide a building level academic performance score for teachers and principals using multiple measures of student academic performance and school assessments. (PDE, 2013a)

Student Achievement Data – indicators of academic achievement, such as: PSSA, Keystone and state value-added assessment calculations.

Summative Evaluation – a teacher performance appraisal designed to provide a record that supports the continuation or the termination of a teacher's employment.

Supervision – The professional coaching of a teacher by the principal.

CHAPTER II

Literature Review

This chapter begins with an overview of the significant role the principal has in the instructional leadership of the school. It goes on to examine the need for change in the teacher evaluation system from the traditional model to the present system. It details the research of Danielson's Framework and makes the case for the need for professional development for the principal.

The Influence of the Principal

Sense of efficacy is the belief in one's own ability to be able to perform a task or achieve a goal. It is not the ability itself, but the belief in one's ability (Leithwood & Jantzi, 2008). While research shows that principals do have an influence on student achievement through the hiring of quality teachers and overseeing instructional quality (Clifford, Behrstock-Sherratt, & Fetters, 2012), do principals have the self-efficacy to believe that they can make a difference? The research of Leithwood and Jantzi (2008) and Wahlstrom, Louis, Leithwood, and Anderson (2010) found weak, but significant, affects of principal efficacy on student learning. Leithwood and Jantzi's (2008b) research found a relationship between principals' efficacy and the proportion of students reaching and exceeding proficiency. The research of Wahlstom, et al. (2010) found there were gains in student learning when using data initiatives, but only when the principal held the belief that improvement was possible.

Kersten and Israel (2005) specifically asked school administrators if they believed they could make a difference in teaching and learning. The results of their research showed that if principals were given the time for increased communication opportunities,

data-driven targeted staff development, teacher peer coaching and mentoring, and the principal modeling specific lessons, instructional techniques, and co-teaching, they would have more of a direct impact on teaching and learning.

The power of the principal to be an instructional leader has been a topic for much research. Strong, positive leadership was found to be a universal characteristic of effective schools (Edmonds, 1982; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Mortimore, 1993; U.S. Congress, 1970). Not only is leadership the catalyst for turning around troubled schools, researchers were unable to discover any troubled schools that were turned around that did not have a highly capable leader (Leithwood et al., 2004). The leader's attention to the quality of instruction makes leadership the number one characteristic of effective schools (Edmonds, 1982). Instructional leadership has been broadly defined as the action taken by the principal (or others designated as instructional leaders) that directly affects the learning of students (DeBevoise, 1982; Leithwood & Jantzi, 1998; Murphy, 1992; Sergiovanni, 2006). In an attempt to conceptualize instructional leadership, scholars have taken the research from studies of effective schools and developed lists of characteristic behaviors of effective leaders.

Conceptualization of leadership. The difficulty in evaluating the principal's role as an instructional leader lies in the need for a clear conceptualization of leadership in order to develop a consistent basis for examining its effectiveness (Mortimore, 1993; Persell, Cookson, & Lyons, 1982; Leithwood & Duke, 1998). Andrews and Soder (1987) developed four broad categories for defining principals as instructional leaders: the principal as resource provider, instructional resource, communicator, and visible presence. In two studies using this conceptualization of the principal's role, Andrews and

Soder (1987) and Smith and Andrews (1989) argued that teachers who perceived their principals to be average or strong leaders in these categories had high student achievement.

Hallinger and Murphy (1986) developed the most fully tested and widely used conceptualization of leadership characteristics in their measurement tool, The Principal Instructional Management Rating Scale (PIMRS). The PIMRS is a self-reporting survey comprised of twelve dimensions of instructional leadership, including framing and communicating goals, supervising and evaluating instruction, coordinating curriculum, monitoring student progress, protecting instructional time, maintaining high visibility, providing incentives for teachers and learning, promoting professional development, and enforcing academic standards. Under these headings are 71 principal behaviors and practices that teachers and principals rate on a Likert scale.

Hallinger (2008) examined 119 studies using the PIMRS from 1982 (when the tool was developed) until 2006. He concluded that the PIMRS is a reliable means of collecting data, although not as valid at the secondary level as it is at the elementary level. It was also discovered that due to methodological limitations, many of the earlier studies were unable to show indirect effects of principals' effectiveness. This became an area of significant, additional research.

Relationships between principals behaviors and student achievement. Early studies searched for a unidirectional relationship between principal behaviors and educational outcomes (Persell, et al., 1982; Hallinger & Heck, 1998). Case studies show that principal leadership is directly related to school effectiveness (Gentilucci & Muto, 2007; Sweeney, 1982) and while there are some recent empirical studies that show direct

relationships between principal leadership and student achievement (Branch, Hanushek & Rivkin, 2012; Silva, 2010), most empirical research studies are unable to show a significant direct relationship between leadership and student achievement (Hallinger & Heck, 1998; Leithwood et al. 2004; Roberson, 2010; Witziers, Bosker, & Kruger (2003).

Pitner's framework (Pitner, 1988) offers several structures, which researchers may use to study the indirect-effects of principals' leadership on student achievement. Using Pitner's framework, Hallinger and Heck (1998) conducted a synthesis of research using the direct-effects model (comparing principal behaviors to student achievement), and the mediated-effects model (where principal effect was examined through the influence on indirect paths including the work of teachers, the culture of the community and events).

The research results of Hallinger and Heck (1998) found that mediated-effects studies show more consistent findings than the direct-effects model. An indirect relationship between principal leadership and school outcomes was found when one study of direct-effects was re-analyzed using a mediated-effects model and the structural equation modeling analysis tool. The challenge of studying the indirect effects of principal leadership to student outcomes lies in determining the mediating variables that leaders influence which will have the most significant impact on students.

More complex data analysis. Using an indirect framework called for the development of more sophisticated methodology tools (Hallinger & Heck 1998, Mortimore, 1993). Research using more robust analysis and software programs to analyze the data now included structured equation modeling (SEM) (Hallinger & Heck, 1998), latent change analysis for longitudinal data (Hallinger & Heck, 2010), hierarchical linear modeling (Anderson, 2008; Marks & Printy, 2003) and path analysis (Leithwood &

Jantzi, 1998; Leithwood & Jantzi, 2008; Sheppard, Hurley, & Dibbon, 2010). While these more complex analytical tools demonstrated a definite relationship between principal leadership and student achievement, correlational research designs do not prove causality (Duke, 1987; Leithwood et al., 2004; Marzano, Waters & McNulty, 2005; Shoemaker & Fraser, 1981; Persell et al., 1982; Glickman, Gordon, & Ross-Gordon, 2001). Yet, educators continue to assume causation and implement programs based on the findings.

Unfortunately, the volume of research has grown so rapidly that traditional analytical approaches to summarizing and synthesizing the findings of these studies in a narrative format is overwhelming. Meta-analysis of quantitative studies allows researchers to synthesize statistical results from groups of studies with the same problem (Gall, Gall, & Borg, 2007). Rather than comparing studies, researchers can compare subsets of studies (Robinson, Lloyd, & Rowe, 2008). Meta-analysis, developed by Gene Glass, is a way to "compare apples to oranges" (Rudner, Glass, Evartt, & Emery, 2002).

Although it is time consuming to research primary sources, the advantages of using meta-analysis are several: the analysis focuses on the magnitude of effect rather than statistical significance, effect size can be applied to any statistic and measure, and the effect size from different studies can be determined and reported as the mean magnitude of effect size across a set of studies (Gall et al., 2007).

Marzano et al. (2005), Witziers et al. (2003), and Robinson et al. (2008), each conducted a meta-analysis on studies examining the relationship of principal leadership on student achievement. Witziers (2003) found little effect (an average correlation of .02) of the principal's leadership in 37 multi-national studies. Marzano (2005), on the

other hand, found principal leadership to have a profound effect (an average correlation of .25) on student achievement. To explain this discrepancy, Witziers (2003) pointed out that international conceptualization and operationalization of educational leadership may differ from that of the American model. The studies in Marzano's research were categorized homogeneously by grade level and outliers were excluded from the study. Marzano (2005) justifies excluding the outliers due to the fact that conceptual and statistical outliers often involve factors that are outside the focus of the analysis. The Witziers' studies were generally focused on direct-effects compared to Marzano who studied indirect-effects. Robertson et al. (2008) align their meta-analysis to Marzano's, in that their study also looked at indirect-effects of leadership and found positive effects of leadership on academic and non-academic achievement. Robertson's analysis is more rigorous as evidenced in the quality of the 27 peer-reviewed published primary sources. The Marzano and Witziers research only contained 20 and 15 published sources respectively.

Measuring outcomes. Most studies comparing principal effectiveness and student achievement measure success using standardized achievement scores. Unfortunately, these scores are a narrow basis for judging outcomes (Duke, 1987; Persell et al., 1982). Basing improvement solely on test scores is comparable to measuring the value of a car solely on the basis of its miles per gallon (Cuban, 1984). If we measure only the lower order skills on standardized tests, we will sacrifice problem-solving skills, creativity and critical thinking (Stedman, 1987). Some researchers have chosen to measure school effectiveness using student engagement (Leithwood & Jantzi, 1998; Leithwood & Jantzi, 1999; Day, Sammons, Hopkins, Leithwood, & Kington, 2008),

teacher morale and enthusiasm for work (Sheppard, Hurley, & Dibbon, 2010) and school improvement capacity (Heck & Hallinger, 2010). It is also understood that context is a crucial variable when determining which behaviors will have the most impact (Persell et al., 1982; Sweeney, 1982; Hallinger, 2003; Robinson et al., 2008; Sheppard et al., 2010) and there is no construct that is valid for every context (Hallinger & Heck, 1998). Schools may be at various stages of reform, have diverse characteristics of culture and have different standards of effectiveness that will require distinctive leadership focus. Context will impact the generalizability of the research findings.

What should researchers measure? Reviews of the literature have shown that principals have an indirect effect on student achievement (Cotton, 2003; Day, et al., 2008; Leithwood et al., 2004; Wahlstrom, et al., 2010). The closer the principal gets to the core of teaching and learning, the more likely he/she is to have a positive impact on student achievement (Gentilucci & Muto, 2007; Robinson et al., 2008). It would seem prudent then to measure the effect that principals have on the effectiveness of teachers as well as a principal's efficacy on improving a teacher's effectiveness.

The Principal's Role as Evaluator

The duality of evaluation. What we measure conveys what we value. Evaluation formalizes the communication of organizational goals, standards for instructional delivery, and values to the school community (Wise, Darling-Hammond, McLaughlin, Bernstein, 1984). Evaluation can be broadly defined as having two assessment purposes: formative and summative. The goals of formative assessment are to provide constructive feedback, to give direction for professional development, and to create a culture of learning for teachers and administrators in their collective efforts to

improve student learning (Haefele, 1993; Ovando, 1994). On the other hand, summative evaluation is used to screen unqualified teachers and provides evidence for termination of incompetent personnel that can withstand the scrutiny of the professional and the judicial communities (Haefele, 1993; Ovando, 1994). Summative evaluation also allows for promotion and recognition of effective teachers.

Danielson and McGreal (2000) summarize the purposes of evaluation as quality assurance (summative) and professional development (formative). Danielson (2011) goes on to say that quality assurance is when a principal can produce proof of the effectiveness of his/her teachers. This requires a consistent definition of good teaching, a common understanding of this definition, and skilled evaluators who are able to recognize evidence of classroom instruction that matches the definition, evaluators who can interpret the specific levels of performance, and evaluators who can engage teachers in productive conversations.

Principals may willingly accept responsibility for their own actions, but they are uncomfortable holding teachers accountable for their instructional performance (McGrath, 2000). By virtue of the fact that the term evaluation is used to represent both summative and formative assessment, there is an ambiguous relationship between the two forms of evaluation due in part because both assessments are delivered by the same evaluator (Danielson & McGreal, 2000; Haefele, 1993; Holland, 2005). Evaluation is a "tug of war" between the evaluator as a judge and the evaluator as a coach (Nolan & Hoover, 2004).

Summative evaluation*:* A critical component of summative evaluation is the removal of ineffective teachers. Painter (2000) describes three difficulties in determining

a teacher's incompetence. First, there is a presumption that every teacher should be perfect or one would believe that the principal has failed. If the reality is that every profession has flawed professionals, then what is the acceptable level of incompetence for a professional? This brings us to the second problem of determining a teacher's competence. There is no precise definition of incompetence. Each principal has his/her own criteria for competence and standard of success. Finally, the context in which a teacher works may contribute to the teacher's success. Some teachers may be more competent working with gifted students, struggling students, or culturally diverse students. What matters is the competency of the teacher in the assignment for which the teacher is being evaluated.

In her research, Painter (2000a) found that principals addressed incompetent teacher performance first by coaching the teacher to a satisfactory standard of instruction, and then either counseled an ineffective teacher to retire or resign, using the evaluation process to remove the teacher, or, to a much lesser extent, facilitated a transfer for the teacher.

Generally, principals are able to identify unsatisfactory teachers (Sartain, Stoelinga, Brown, 2011). Research studies, though, have found that approximately 1% of teachers receive less than satisfactory ratings on their summative evaluations (Frase, 1992; McGrath, 2000; Weisberg, et al., 2009), yet principals and teachers believe that teachers are not functioning at the level the ratings would suggest. Frase (1992) found that lessons of satisfactory teachers were focused on low-level cognition, with extensive drill and practice, resulting in limited student to student and student to teacher

interactions. Weisberg, et al. (2009) uncovered that teachers report receiving little or no meaningful feedback from their evaluations.

In summaries of the research (Danielson, 2000; Donaldson, 2009; Ellett & Teddlie, 2003), it was found that teacher evaluation has not substantially improved instruction. A significant proportion of a principal's time is spent in the process of teacher evaluation, yet principals believe that there is limited impact on teaching and learning (Kersten & Israel, 2005). Evaluators give satisfactory evaluations as a motivator in order to build trust and support of teachers. Unfortunately, motivation is used at the expense of critical feedback (Donaldson, 2009).

The traditional evaluation system leads to a "culture of passivity and protection" where the teachers don't question how they can improve their instruction and the principals give satisfactory ratings (Danielson & McGreal, 2000). Teacher evaluation systems are broken and must be reformed in order to improve the instruction of students and the quality of the teacher (Donaldson, 2009).

Formative evaluation. After interviewing 3000 educators regarding evaluation, Marzano et al. (2005) found that 76% of educators stated that measurement and development were the dual purpose of evaluation, but that development should be the dominant factor. Unfortunately, the research of Weisberg, et al. (2009) found that, even though teachers expressed a strong desire for more concrete, detailed feedback to promote their professional development, generally educators do not provide high-quality feedback for teachers following observations.

Relying on conventional classroom observations to evaluate teachers is ineffective, inefficient and does not help teachers grow professionally (Howard &

McColskey, 2001). Principals observe less than 0.1% of a teacher's annual instruction in an over-glamourized lesson where teachers take minimal risks and rarely admit to problems. The top-down communication does not support collaboration and instead reinforces isolation and shuts down teacher learning (Marshall, 2005). Limited administrator training, poorly defined descriptors of best practices, and a lack of a specific and descriptive rating scale make for a flawed system of evaluation (Danielson & McGreal, 2000).

Danielson (2011) and Sawyer (2001) identified several flaws in the traditional evaluation system identified by educators. First, the evaluation system is time intensive for administrators. It is an annual routine that becomes a nuisance for principals and teachers. In addition, the evaluation criteria are outmoded and there is a need for the criteria to reflect current research-based best teaching practices. Principal directed conversations offer simplistic evaluation comments giving little new or challenging information to the teacher for improvement. There is a lack of inter-rater reliability between evaluators and lastly, most attention is given to the novice or unsatisfactory teachers with little coaching for the career professionals.

Time for change. Elmore (2000) posits that if we put instructional improvement at the center of our theory of leadership, then all other skills must be defined as instrumental to this goal. To bring about comprehensive change, teacher evaluation systems should not just focus on the best and worst teachers, but should focus on improving the instructional effectiveness of all teachers (Papay, 2012; Shakman, Breslow, Kichanek, Riordan, & Haferd, 2012). Authentic leadership must do away with the bureaucratic process of evaluation and concentrate on the shared values and moral

purposes of improving instruction (Sergiovanni, 2000). The core of achieving instructional improvement is in creating a supportive culture where teacher growth is valued (Colby, Bradshaw, & Joyner, 2002).

Marshall (2005) fine-tunes the concept that the enhancement of a teacher's effectiveness will improve student achievement. He proposes a new theory of action behind formative and summative assessments suggesting that principals move away from owning the evaluation system and give ownership to teacher teams to work collaboratively. Research shows that teachers value and benefit from collaborating in the analysis of data and in collective participation in professional development activities creating cultures focused on teaching and learning (Colby, et al., 2002; Desimone, Porter, Garet, Yoon, & Birman, 2002; Holifield & Cline, 1997). Teacher evaluation should be a tool to build the instructional capacity of teachers (Papay, 2012).

Standards-based tools. In 1988 the Joint Committee on Standards for Educational Evaluation developed Personnel Evaluation Standards. Holland (2005) proposes an additional six standards that would make the process more meaningful for educators as well as meet the accountability expectations of legislators. These standards are: 1. Differentiated procedures to align with the respective levels of teacher experience and professional development needs, 2. Teachers and administrators should work collaboratively to establish goals and analyze data to justify the achievement of the goals, 3. Evaluators of teachers should use data from multiple sources, from multiple points in time, provided by multiple evaluators, 4. Evaluations should be formative and summative with the majority of resources used in the formative assessment processes, 5. Evaluation of teachers should be tied to the teacher's professional development goals as

well as the school improvement goals, and 6. Evaluation policies should be well defined, articulated and clearly communicated.

These standards have been accepted by many of the leading authorities who have attempted to create a framework for a fair and comprehensive evaluation system for classroom observations. Marshall (2005) proposes a twelve-step program that includes an acronym simplistically representing the characteristics essential in every classroom. The umbrella SOTEL represents physical and psychological safety of students, clear objectives, effective teaching, and engaged students who are learning. Although the twelve steps are aligned with Holland's standards, they are not as clearly defined as teachers and administrators need.

Marzano et al. (2005) synthesized their research into three primary characteristics of teacher evaluation systems designed to improve instruction: a system that is comprehensive and specific, one that provides a developmental scale, and one that acknowledges and rewards growth. Marzano's research identified forty-one classroom strategies and teacher behaviors that support student achievement. To guide a teacher's skill development, he has developed a scale ranging from "not using" to "innovating" to represent the level of proficiency of the teacher on each strategy or behavior. At the end of the year, the teacher is evaluated on a pre-set improvement goal with two scores. The first score is placed on the rubric with the descriptors: "not using", "developing", or "innovating", and the second score shows growth using the descriptors: "advanced", "proficient", and "failing."

Danielson and McGreal (2000) and Papay (2012) believe that an evaluation system that supports teachers' growth and development fills both requirements for

accountability as a measurement tool and reflective practice that gives targeted feedback to drive continued instructional improvement. The requirements for this type of summative evaluation based on standards are: a clear evaluative criteria, the citing and weighing of evidence against a rubric that instruction is meeting that criteria, a set of procedures for collecting multiple forms of data regarding a teacher's performance, neutralizing of bias, and a non-judgmental demeanor on behalf of the evaluator (Danielson & McGreal, 2000; Odden, 2004; Papay, 2012). One of the most comprehensive and widely used standards-based teacher evaluation tools used is The Framework for Teaching (Danielson, 2007).

The Framework domains and components reflect current views of "best practice" in instruction. The levels of performance describe teaching practice that is active, differentiated, inclusive, engaging for students as they participate in meaningful activities, and cultivating a community of learners (Danielson, 2007). A paramount component of the system includes teacher reflection. This standards-based evaluation system appears to have the potential to provide measurements of teacher practice that is strongly related to student achievement (Milanowski, 2004).

Validity and reliability of standards-based evaluations. Validity is the extent that an assessment measures what it infers it will measure (Gall, et al., 2007). In estimating the validity of standards-based observation evaluation supporting effective teaching, we would want to measure the success of the instruction against the learning of the students. Validity is one of the most important factors in the success of standards-based evaluation, but it is the most difficult to assess (Papay, 2012) as will be shown in the following research.

Reliability refers to the consistency of an assessment score (Gall, et al., 2007). In other words, if a teacher were observed a second time, whether observed by the same or a different observer, the scores would be the same. Achieving inter-rater reliability in observations is difficult as two different evaluators have the potential to interpret the evidence of instruction differently (Papay, 2012). Evaluators need to be calibrated to a standard (Shakman, 2012). This can be achieved through professional development on the criteria for evaluation, and participating in norming exercises (Shakman, 2012). Danielson (2007) and Donaldson (2009) support comprehensive training to ensure evaluators interpret the standards in a consistent manner. Another way to ensure reliability is to have multiple evaluators observe the same teacher.

Establishing the validity and reliability of the Framework and the case for professional development for principals. Believed to be the largest study ever to investigate the relationship between multiple evaluation instruments, the MET Project sponsored by the Bill and Melinda Gates Foundation, was a three-year study that included the analysis of the validity and reliability of the Framework for Teaching. Three thousand teachers representing seven districts from across the country provided videos for evaluators to score. Using the Framework, teachers were given scores on their instructional strategies by evaluators. To determine the validity of the Framework to accurately measure the effectiveness of the teachers, these scores were correlated to student achievement scores. The key to proving the validity of the Framework was in the random assignment of the students to teachers (Kane, McCaffrey, Miller, & Staiger, 2013).

The research was able to show that teachers identified as most effective as assessed by the Framework, also produced the greatest achievement growth in their students. Two limitations were noted regarding the randomization of the student assignments. First, random assignment could not be achieved across different school sites. The evidence from the study does not inform the between school comparison. Second, because the assignments were made early in the summer before students and teachers assignments were certain, the compliance of keeping randomized students in the same class for the entire year was compromised. Nevertheless, in order to take advantage of the effect of the actual teacher, a statistical approach called, "instrumental variables" was used to infer the impact of the actual teacher (Kane, et al. 2013).

Assigning 129 observers, all principals and teachers, to observe and score actual teacher lessons determined the reliability of using the Framework. Each video lesson observed received twenty-four scores. It was found that administrators rated their home-school teachers slightly higher than outside evaluators, although the difference was not significant enough to change the rankings due to favoritism bias or prior impressions. The reliability of using school personnel was .65 or above. When evaluators viewed more than one lesson of the same teacher, it was found that positive and negative first impressions lingered. It was also found that an accurate observation rating requires two or more lessons evaluated by different observers (Ho & Kane, 2013).

This research determined that a well-designed evaluation tool could provide reliable feedback on classroom observations that points out the strengths and areas for growth for a teacher. The tool can be used to identify aspects of teaching that will predict

student learning. Each measure should provide teachers with accurate and meaningful feedback (Ho & Kane, 2013).

While the MET Project budgeted $40 million for the implementation of the new teacher effectiveness system in Pittsburg, PA (Bill and Melinda Gates Foundation, 2009), only three hours of training was given to principals and teachers before implementing the program. In a follow-up survey, principals reported not being prepared to implement the program and were looking for continuous learning opportunities for the process as well as in the follow-up with the teacher (Lane & Horner, 2011).

Modeled after the MET Project, Team PA conducted a pilot study for a new evaluation system in Pennsylvania. Partners included in Phase 1 of the pilot were leaders from Pennsylvania Department of Education (PDE), Pennsylvania State Education Association (PSEA), school districts, intermediate units, policy makers, and members of the business community. In order to design a teacher evaluation system that improves instruction and promotes student learning, input from teaching professionals was deemed to be critical (Donaldson & Donaldson, 2012). Teachers involved in the process through the PSEA would become the cheerleaders in promoting the new system.

The goal of Phase 1 of the pilot was to design the evaluation system for teachers and principals. Participating in the pilot program were: Allentown, Cornell, and Mohawk Area school districts, and Northwest Tri-County Intermediate Unit 5. Charlotte Danielson's *Framework for Teaching* (Danielson, 2007) was used as the instrument for classroom observation. The decision to use Danielson's framework was a result of an extensive research study, the MET project, in which Danielson's framework was studied for its reliability and validity. A major finding of the MET study was that teachers who

demonstrated the types of practices emphasized within Danielson's framework showed greater student gains than teachers who did not score as well using the evaluation protocol (Bill and Melinda Gates Foundation, 2012).

Participants in Phase 1 of the pilot were given three hours of training and five months to implement the evaluation system. Several methods were used to gather information about the teacher evaluation pilot from participants including surveys, focus groups with teachers, and interviews with principals. Of the 160 teachers involved in the pilot, 71% participated on the surveys and 63% of the 30 principals participated. Thirty-four teachers participated in the focus groups and six principal evaluators were interviewed. The findings showed that overall the evaluation rubrics supported good teaching and the participants were responsive to the process. Teachers recommended that they have more training sessions over a longer period of time and in multiple sessions. They wanted the trainings to be more content and grade level specific. Teachers wanted more specificity in the rubric; specifically the requested types of evidence for each of the domains so they could delineate proficient from distinguished ratings. To complement the evidence from the observational data, teachers would like to include portfolios, artifacts, and their preconference forms (Lane & Horner, 2011).

Principals reported that the training did not adequately prepare them to use the Framework to evaluate teachers. They, too, were looking for more intensive, comprehensive training over a longer period of time to clarify the vagueness of the procedures. Also, principals were looking for more guidance in regard to the nature of evidence for each competency and in goal setting for teachers (Lane & Horner, 2011).

Phase 2 of the pilot was delegated to Mathematica Policy Research, which used student-level data to develop Value-Added Models (VAM) for estimating the effects of educators on student growth. [VAM does not measure student growth, but is an estimate of an educator's contribution to student growth (Lipscomb, Chiang, & Gill, 2012).] Decades ago, there was no capacity to collect, analyze, and track student growth. Today with current computer software and statistical models based on research, psychometricians can now correlate the learning gains of a student to the teacher (Elmore, 2000). Research conducted by Sanders and Rivers (1996) concluded that longitudinally linked teacher-student data was able to distinguish more effective teachers from less effective teachers.

Two researchers, Gallagher (2004) and Milanowski (2004) also studied the validity of using the Framework as an observational tool to increase student achievement. Both were considered high stakes testing as the results were to be used to validate the use of the evaluation in determining performance pay.

Gallagher's (2004) research used an adapted form of the Framework and his sample had a unique profile of one elementary charter school in Los Angeles with 100% Title I and free and reduced lunch, and 85% English language learners. His results showed a positive, strongly significant relationship between teacher evaluation scores and student achievement in reading. Although there was a positive relationship in math, the scores were not significant. Through the qualitative portion of his research, he found that literacy expertise was the focus of the most recent staff development. Teachers and the administration had a clear vision of effective reading instruction. This finding adds to the research that strong content knowledge impacts effective teaching. It also reinforces the

concept that teacher and principal efficacy play a role in a teacher's and principal's ability to effectively teach and effectively evaluate when they are knowledgeable about the subject.

Milanowski (2004), whose research took place in Cincinnati public schools, used the evaluation of six classroom observations, four by teacher observers and two by the principal, along with teacher portfolios for 212 teachers to compare with student achievement on state and district tests in reading, math and science. The students participating in the study were representative of the total student population. Cincinnati made a substantial investment in professional development and only those evaluators who met the standard were allowed to evaluate the instruction.

Milanowski (2004) found correlations of .27 in science, .32 in reading and .43 in math. The conclusions from this research substantiated the MET results that the scores of a rigorous teacher evaluation system can substantially relate to student achievement. One limitation of the research was that the teachers didn't always teach in the tested grade levels or teach the subjects tested; therefore, there was a small sample of teachers who actually met the criteria for comparison.

A more recent study in Cincinnati conducted by Taylor and Tyler (2011) confirmed Milanowski's (2004) research regarding math achievement. This study found that mid-career teachers improved their effectiveness following their evaluations using the Framework with the least skilled teachers benefitting the most. The researchers pointed out that their results may not generalize to other districts that have not made the substantial investment in their teacher evaluation system as Cincinnati has made. The intensive training program in Cincinnati ensures quality feedback to their teachers.

In a two-year study of Chicago's Excellence in Teaching Pilot, Sartain, Stoelinga, and Brown (2011) randomly selected schools and the teachers within these schools for their study to ensure that the information gleaned from their research could be generalized across the city. They conducted 955 observations of 150 teachers to gather quantitative data regarding the use of the Framework to evaluate teachers. A statistical model was used to compare the evaluation results to value-added scores from standardized testing in math and reading. Results of the research showed that the teachers who were rated highest using the Framework, had the highest student achievement. Similar, though, to the limitation that Milanowski (2004) found, Sartain, et al. (2011) also found that many teachers did not teach a subject or grade that was assessed. In addition, Sartain et al. (2011) found, due to the support of team teaching, it was difficult to assign students to specific teachers.

To show reliability of the instrument, 499 observations were conducted with 257 teachers. Principals and outside evaluators watched the same video observations but scored them separately. Eleven percent of principals scored lower than the outside evaluators and 17% scored higher. Overall, though, principals and outside evaluators scored the same with the same rating.

What was gleaned from the qualitative portion of the research (interviews of 39 principals and 26 teachers) was that while conferences were more focused on instructional practice and improvement, many principals lack the coaching skills required to have deep, meaningful conversations about teaching practices. While teachers were positive about the implementation of the Framework, they were negative about how it was used by unprepared principals. Effective evaluation will require principals to make a

shift in their mind sets from the idea that principals "just know" good practice, to evidence based judgment that allows principals to diagnose a teacher's strengths and areas for improvement in the teacher's instruction. It is imperative that principals participate in professional development to re-conceptualize for themselves this new evaluation system (Sartain, et al., 2011).

Sawyer (2001) summarizes a two-year action research project initiated in Washoe County School District, Nevada where 60 principals and 1765 teachers field-tested the Danielson Framework as their new evaluation system. The focus of the project was to determine whether the time allotted to goal setting was worthwhile. Using surveys and focus groups to gather data, it was found that the time spent in goal setting increased meaningful conversations between the teachers and principals. Veteran teachers were revitalized by the specificity of the feedback and novice teachers were appreciative of the explicit expectations and indicators of success. Like Sartain, the negative comments about the process were from principals who did not have adequate training in using the reporting documents and in identifying evidence to support teacher behaviors.

Kimball, White, Milanowski, and Borman (2004) expanded the work of Sawyer (2001) in Washoe County School District and analyzed the relationship between teacher behavior, as measured through the Framework, and the amount of student achievement attributable to teachers. Teachers who had evaluation scores and could be matched with students who had pre and post test scores were included in the study. The results were mixed providing only tentative evidence for validity of the evaluation system. While the relationship was positive between scores, the scores were not statistically significant for every grade level.

There were several reasons this research did not show the degree of significance as previously discussed research. First, the evaluation cycle for this district is one where teachers are evaluated on different domains each year. Post-probationary teachers move into a cycle of evaluation that does not include all components every year. These teachers were evaluated using a supplemental evaluation form. A performance composite was used to measure key elements of the domains. Only seven of the 23 components were measured. Second, because this district follows a year-round school schedule, teachers had a different number of instructional days. Third, there was a question regarding the reliability of the tests compared to what was actually taught. This may have been a confounding factor. Finally, a common theme of research limitation is that there was limited emphasis on evaluator training and inter-rater reliability.

In summary of the research, the preponderance of evidence shows the Framework to be a valid and reliable means to evaluate teachers, yet it points to the essential need for professional development for principals. To maximize on the success of standards-based evaluation systems, principals should be sufficiently trained in the use of the tools, ratings and systems of reporting teacher effectiveness (Danielson, 2007; Kimball et al., 2004; Marshall, 2005; Sartain, et al., 2011; Sawyer, 2001), the knowledge about effective teaching practices as defined by the standards (Danielson & McGreal, 2000; Papay, 2012; Sawyer, 2001; Wise, Darling-Hammond, Mclaughlin, & Bernstein, 1984), and in conducting critical conversations with teachers to improve practice (Colby et al., 2002; Donaldson, 2009; Marshall, 2005; Pajak, 1990; Sartain, et al., 2011; Taylor & Tyler, 2011).

Professional Development for Principals

Guiding principles. Traditionally, professional development consisted of one-day workshops that lasted a few hours where the participant was a passive listener. While the participant may come away with a few practical tips, the sessions were usually in isolation having no connection with the reality of a participant's needs (Shakman, 2012). There were seldom follow-up sessions to evaluate the learning or implementation of the learning and subsequent workshops may have focused on a completely unrelated topic (Corcoran, 1995).

With heightened accountability there should be heightened support for professional development of principals (Shakman, 2012). The laissez-faire approach to staff development must be transformed into purposeful, productive programs directly related to the work of principals (Corcoran, 1995). Professional development should model constructivist teaching (Corcoran, 1995; Danielson & McGreal, 2000). Danielson and McGreal (2000) explain that one of the principles of adult learning theory is that when one develops professional learning as a result of self-assessment and self-directed inquiry, the learning is more meaningful and is more likely to be sustained than when professional development is imposed by outsiders. Professional learning should begin with thoughtful reflection (Barth, 1986). It is not experience that teaches, but the reflection on the experience that enables us to analyze our work, clarify our thinking and support in-depth learning (Barth, 1986; Corcoran, 1995, Danielson & McGreal, 2000; Desimone, et al., 2002).

Professional development should be linked to the specific needs of the principal and the community the principal serves (Bennett, 2002; Danielson & McGreal, 2002).

The closer the professional development is linked to the site-based needs, the greater the impact will be (Corcoran, 1995). In his guiding principles for professional development, Bennett (2002) includes holding everyone to high expectations and harnessing the power of data.

Two of the most widely proposed guiding principles of professional development are creating a culture of learning and collective participation. Peterson (2002) proposes that successful professional development for principals requires a strong, positive culture identified by professional relationships built on trust and camaraderie. Survey results from Pajak (1990) highlight the notion that leaders should take time to know something about each individual in the organization. Communicating on a personal level is just as important as communication on a professional level.

In developing a culture of learning, Frase (1992) recommends a paradigm shift from "Status quo is okay" to "Together, we can do better." The foundation of this premise is that everyone has room for improvement, constructive feedback is healthy, success and power lie in intrinsic motivation, and the belief that teachers want to make a meaningful contribution to their students' learning. Danielson and McGreal (2000) believe that there should be no apprehension in a community of learners. Instead, professional development should be intellectually, socially and emotionally engaging.

In a study conducted by the Wallace Foundation, a strong correlation was found between schools with high levels of student achievement and high ratings by teachers of principals who supported a culture of continual professional development (Wahlsrom, et al., 2010). This focus on learning is promoted by allocating resources to staff

development, training the staff in the use of the evaluation tool, and differentiating the staff development (Shakman, 2012).

Collective participation allows colleagues to critically examine the new standards, share and interpret data, and to work and solve problems together (Barth, 1986; Colby, et al., 2002; Corcoran, 1995; Jensen & Moller, 2013; Peterson, 2002). Danielson and McGreal (2000) state that professionals are more apt to consciously reflect on their work if they are collaborating. They also point out that collaboration offers the opportunity to hear an alternate point of view. By listening to others, we are able to give coherence to what we already know (Barth, 1986).

Finally, a guiding principle of professional development is that the training should have an extended duration. Time must be allotted for participation, implementation, and follow-up (Corcoran, 1995; Danielson & McGreal, 2000; Desimone, et al., 2002; Peterson, 2002). "Paradoxically, professional development can be energy and time depleting and energy and time replenishing" (Barth, 1986, p. 93). It is important that we ensure both facets of this paradox, especially the replenishing aspect.

Background for principal professional development. The National Center for Education Statistics reports that only 75.2% of districts nationwide reported offering professional development in evaluation and supervision to principals in 2003-2004 (Strizek, Pittsonberger, Riordan, Lyter, & Orlofsky, 2006). Despite the critical role of evaluation training, it is not universal (Donaldson, 2009). Elmore (2000) states that leaders must be able to model the learning that they expect from others and that the strength of leadership is not in the formal dictates of effectiveness expectations, but in the required knowledge and expertise leaders must possess to support improvement.

As the emerging role of the principal focuses on instructional leadership, leadership preparation programs have not prepared principals to tackle the tasks associated with these responsibilities (Elmore, 2000; Lashway, 2003; Levine, 2005; Portin, Schneider, DeArmand, & Gundlach, 2003). In Levine's four-year study of America's education schools (2005), it was noted that 47% of principals and 39% of all administrator alumni characterized the curriculum of their leadership preparation programs as outdated and ineffective. Principals believe that experience has been the best teacher and that most of what they have learned, they have learned "on the job" (Portin, et al., 2003).

Educators and policymakers are recognizing that front-loading principal training and following up with sporadic, isolated professional development opportunities is insufficient in producing quality leadership (Lashway, 2003). Even providing detailed rubrics and rating scales is not enough to ensure qualified evaluators (Kimball, et al., 2004). In the study conducted by Kimball et al. (2004) it was found that principal training was "front-ended" regarding the implementation of the Framework. Training focused on interpreting the rubric, the domains, and sources of evidence, but there was no training in inter-rater reliability or sustained staff development. When principals were asked about their weaknesses in evaluating teachers, they identified their lack of ongoing, personnel evaluation training.

Leadership training cannot be a single event, but continuous learning must be promoted if principals are to improve the practice and performance of teachers (Elmore, 2000; Kimball, et al., 2004; Lashway, 2003, & Portin, et al., 2003). Many states are mandating a "second-level" certification that requires formal mentoring, reflection,

portfolio development, and performance assessments (Lashway, 2003; Shakman, 2012). At this time, these certification programs are voluntary and not linked to state licensing (Lashway, 2003). In order to improve the consistency of evaluators in collecting of evidence and evaluating of observations, PDE is recommending that administrators enroll and pass an inter-rater reliability program, Teachscape FOCUS.

Training in the use of tools, ratings, and systems of reporting. Evaluators making judgments regarding teacher effectiveness must be trained in order to create conclusions that are accurate, reliable, and based on evidence (Danielson & McGreal, 2000). The success of the Teacher Observation and Practice component of the new Pennsylvania teacher evaluation system depends on the proficiency of the observer (Teachscape, 2013). Teachscape FOCUS (formerly Framework for Teaching Proficiency System) is PDE's endeavor to provide training for principals as they implement the new Framework for Teaching. The emphasis of this online professional development is to define effective teaching and best practices that will increase inter-rater reliability in order to promote consistency and fairness for the evaluation process (PDE, 2013a). It is an opportunity for principals to refine their skills as they collect evidence during classroom observations that correlates to the Framework and then use this evidence to provide feedback to teachers to foster their growth. This process will help principals to determine the overall performance rating of a teacher in each of the domains (Volkman, 2013).

Based on the 2011 version of the Framework for Teaching, the training will deepen the understanding of Domains 2 and 3, which have been condensed to four components each. The online, self-paced instruction requires 20 – 25 hours of study in

three sections: Observer Training, Scorer Practice, and the Proficiency Test. Observer Training consists of a *Minimizing Bias* module, which will help principals to identify relevant evidence, bias, and interpretation. Following this module, there are eight component modules that delve deeper into the understanding of the Framework. To finish this section, there is a module on *Applying the Framework* where principals will begin to relate the evidence to all eight components of the Framework. Scorer Practice provides five videos that principals will score. These results will be compared with the scores of expert scorers and principals will be able to access the experts' rationale for their score (Teachscape, 2013).

The final section of training is the Proficiency Test, which assesses five sub-skills relative to conducting successful classroom observations:

1. Distinguish between appropriate evidence and interpretation
2. Distinguish between appropriate evidence and statements that are biased or suggest professional preferences
3. Recognize evidence that has been mis-categorized to the wrong component.
4. Assign an accurate score for each of 8 components based on a set of evidence
5. Assign evidence to the appropriate component (Teachscape, 2013).

Stage 1 of the test consists of multiple choice items and video "testlets" that are scored on the eight components in Domains two and three of the Framework. Stage 2 consists of three "testlets" targeting content knowledge about specific dimensions of the Framework as applied in the scoring videos. Principals may not move on to the second stage without first passing Stage 1. Two attempts are given to pass each stage (Teachscape, 2013).

PDE is recommending that all administrators apply for licenses and pass this inter-rater reliability program (PDE, 2013a).

Charlotte Danielson (Teachscape, 2013) professes that the success of any evaluation system to be considered reliable and defensible depends on two trained observers agreeing on the evaluation of a lesson. She goes on to say that the goal of this training is for evaluators to be able to take clear notes that are not contaminated by their biases and preferences, organize the notes by components, and then reference the rubric in order to make a clinical evaluation on the level of performance represented by the data. The ultimate goal of this program is for principals to have substantive and meaningful conversations with teachers.

Knowledge of effective teaching practices. Case studies found that the credibility of principals is a critical factor in the success of the evaluation system (Colby, et al., 2002). Leithwood (2005) postulates that if school leaders are going to make significant contributions to student learning, it will require an extensive breadth and depth of knowledge. The most important characteristic of a leader is that the leader is a learner (Barth, 1985).

Marzano, et al. (2005) describe highly effective administrators as possessing extensive knowledge about effective instructional practices. The most successful principals are able to provide guidance to teachers regarding improvement of classroom practices. It is the principal's role to know what kind of professional development a teacher requires in order to improve the delivery of instruction (Fink and Resnick, 2001).

When evaluating, the principal is expected to be the pedagogical and content expert with the ability to make summative judgments and offer suggestions for

improvement. Having pedagogical and content knowledge in all areas is not a realistic expectation. When the evaluator is perceived more as the supervisor, the teacher and principal are able to share their expertise and balance their skills in order to make critical judgments.

Principals Conducting Critical Conversations

The purpose of the post-observation conference is to enhance instruction and student learning (Frase, 1992; Ovando, 2006). The information shared with teachers must be honest and relevant in order for the teacher to modify the instruction, correct errors, or engage in professional development (Ovando, 2006). No longer should observation and feedback be synonymous with evaluation (Danielson and McGreal, 2000; Howard & McColskey, 2001). Observation and feedback should be formative in nature with the focus on reflection of teaching and learning (Danielson & McGreal, 2000; Donaldson, 2009).

Few administrator preparation programs train new administrators in their credentialing programs in conferencing techniques (Haefele, 1993; Ovando, 2006). In their study of Chicago schools, Sartain et al., (2011) found that reflective or post observation conversations were difficult for teachers and principals. They used two ways to assess the quality of post conferencing conversations. First, they measured the type of questions principals posed to teachers and found only 10% of these questions to be of a high level based on Danielson's definition of questioning. They also measured the proportion of time principals and teachers drove the conversation. It was found that the principal controlled the conversation 75% of the time, leaving the teacher only twenty-five percent of the time to explain his/her point of view. Fifty percent of the principals in

this study expressed a desire for more coaching in the type of feedback that would help improve instruction.

McGrath (2000) brings to our attention the strong response teachers have to less than positive feedback. Principals who are not comfortable or confident in providing constructive feedback suffer from "inarticulitus," caused by the fear of the reaction of others (McGrath, 2000). Because of lack of experience or training or because of a previously uncomfortable conferencing experience, principals may fear giving feedback (Seashore Louis, Leithwood, Wahlstrom, Anderson, 2010). The substance of the constructive conversation can be lost in the equivocation of the evaluator. The feedback then is not constructive or sufficient for instructional improvement or to motivate a mediocre teacher (Frase, 1992; Howard & McColskey, 2001; Marshall, 2005; McGrath, 2000). While principals hold themselves to a high standard of accountability and responsibility for their own actions, they struggle with holding teachers responsible for their performance (McGrath, 2000).

When the feedback is perceived as being delivered in a demeaning, sarcastic, or mean-spirited way, even if the advice is accurate, the message is lost on the teacher and there will be no instructional improvement (Frase, 1992; McGrath, 2000). The research of Sartain et al. (2011) found that teachers who had negative interpretations of their conferences with principals, were also skeptical about the ability of the principal to use the evaluation tool accurately and fairly. It is clear that principals are in need of professional development in conducting critical conversations with teachers.

Elements of effective conferencing. One of the "universal building blocks" that affects instruction and student learning is high quality conversations facilitated by a

skilled evaluator that are understood by the teacher (Saphier, 2011). Danielson and McGreal (2000) recommend that districts train principals in observation and conferencing skills in order to make these skills as reliable as possible. Danielson (2011) goes on to say that it is crucial for administrators to schedule observations and conferences into their day so as not to yield to the temptation of letting managerial duties overshadow instructional improvement.

Post observation conferences should be a time for reflection, constructive feedback, and reinforcement (Danielson and McGreal, 2000). Evaluators must reduce the negative mindset associated with feedback by keeping comments as impersonal as possible (Ovando, 1994). Language can also impact the perception of the message. The use of "but" can negate any positive comments that precede the message. Comments should reflect clinical observations rather than be judgmental or biased statements (Danielson & McGreal, 2000; McGrath, 2000). Principals should offer support and optimism (Ovando, 1994).

Setting a goal: The Framework provides clear and specific descriptions of best practices that can be used to develop goals for teachers. Before entering into an observation, establish a narrow focus or goal the teacher is trying to achieve (Danielson & McGreal, 2000; Wiggins, 2012). This should be something relevant and tailored to the individual teacher (Ovando, 1994).

Collecting and analyzing the data: Collect tangible evidence with which to judge the achievement of the goal (Danielson & McGreal, 2000; Ovando, 1994). When giving feedback, effective principals "hold up a mirror" for teachers to see just what occurred in the classroom (Blase & Blase, 2000). Wiggins (2012) suggests that teachers

videotape their lessons once a month to get a bird's eye view of their actual performance and the students' response to the instruction. Analyze the data to determine the strengths and areas of need in the instruction (Ovando, 1994).

Actionable feedback: Explicit feedback opens the conversation for where improvement needs to take place. Rather than using a generic statement or judgmental statement, such as, "The students were disruptive during instruction," clinical evidence such as, "Three different students called out during the direct instruction, for a total of ten interruptions," gives teachers feedback that can be analyzed. This will open the discussion on classroom management strategies. Clinical feedback is specific, neutral, and takes the emotion out of the conference (Danielson & McGreal, 2000; Wiggins, 2012).

Helpful guidance*:* Provide a position of problem-solving (Blase & Blase, 1999). Focus on one or two areas for refinement, being careful not to be overwhelming. Clarify the issue, being sure the teacher has an understanding of what needs to be refined (Wiggins, 2012). Saphier (2011) supports content analysis conversations that enables teachers to frame specifically what they want students to know and what potholes the students may face. Work with the teacher to foster growth, not as a disciplinary procedure, but as a way to improve the performance (McGrath, 2000). Feedback should be linked to professional development opportunities (Orvando, 2006; Saphier, 2011). While Ovando (1994) suggests that feedback should be immediate, Wiggins (2012) prefers the term "timely." During the lesson may not be the time to give feedback, but delivering feedback as soon as possible allows the teacher to implement revisions as quickly as possible.

Progress reporting*:* Ovando (2006) encourages evaluators to provide on-going support, acknowledging achievements toward goals. The "feedback loop" gives teachers a chance to reshape their performance (Wiggins, 2012). Rather than "following up" which implies accountability on the completion of a task, Ovando (2006) prefers to "follow-on" which assures continuing work with teachers as they respond to the plan established at the conference.

Confidentiality*:* In order to create a relationship based on respect and trust, post conference conversations should be confidential. Praise openly, but when having a critical conversation, keep the contents of the discussion private. This will promote cooperation and growth from the teacher.

Principals must rely on the belief that teachers want to do a good job and make a significant contribution to their students' lives (Frase, 1992, McGrath, 2000). Effective conferencing helps teachers to fulfill their intentions by refining their craft. It is the goal of effective conferencing that teachers will increase their reflection, innovation, instructional strategies, risk-taking, planning for instruction, motivation, efficacy, sense of security, and self-esteem (Blase & Blase, 1999).

Organizational and Personal Characteristics of Principals

There is a dearth of research associated with the organizational demographics and personal characteristics of principals in relation to their need for professional development in the process of teacher evaluation. Leithwood and Jantzi (2008) found that a leader's efficacy effects were significantly moderated by the organizational demographics of school size, but there was no significant effect of gender or experience on a principal's efficacy. Bryant (2011) on the other hand found a correlation between

the experience of administrators and their comfort in the implementation of the evaluation process. The more experience an administrator has the more effectively the administrator performs the behaviors associated with the teacher evaluation process. There were no significant relationships between age and gender and a principal's efficacy on teacher evaluation.

Kerrins and Cushing (2000) found that both novice and expert evaluators have a similar knowledge base regarding instructional pedagogy and the management of the classroom, although their application of the knowledge when conducting observations and conferences is quite different. The novice administrators are able to describe the observation, whereas the more experienced administrators are able to interpret the evidence in the instruction in terms of student learning. It was found that novice administrators were unable to evaluate the observation in order to make recommendations for improvement. The implications of these findings suggest that a peer-coaching model of professional development for novice administrators would be beneficial. The more experienced administrator paired with a novice would assist the novice in giving more sophisticated feedback.

Formats for Principal Professional Development

As the role of principal shifts to that of instructional leader and the need for training becomes imperative, the format used to train principals comes in question. The goal of administrator professional development is to increase the professional capacity of the school leadership (Grissom & Harrington, 2010). There is limited empirical research in the efficiency of principal professional development beyond self-reported experiences and perceptions of participants (Davis, Darling-Hammond, LaPointe & Myerson, 2005).

Principals rated university course work for professional improvement less effective in improving professional performance (Davis et al., 2005; Grissom & Harrington, 2010; Levine, 2005). While Davis et al. (2005) contend that a strong collaboration between the university and school districts would allow for on-site training where theory and practical implementation can be merged, this does not always translate into success. Too often instructors consist of part-time faculty or local superintendents and principals with no scholarly expertise of up-to-date research telling "war stories." Conversely, full-time faculty members are often charged with being disconnected from current practice (Levine, 2005). A survey by Levine (2005) of alumni from administration programs found the following areas to be the most critical areas in need of improvement: faculty with more experience as practitioners (56%), more relevant curriculum (40%), upgraded technology (36%), and opportunities for more clinical experiences (35%).

In a study by Grissom and Harrington (2010), strong evidence was found that not all modes of professional development for principals were equally effective in improving principal performance. Looking at several forms of continuing professional development, this study found a significant positive association between principal participation and the effectiveness of formal mentoring and coaching.

Formal mentoring is characterized as an experienced administrator working closely with a novice principal to solve problems and work through the day-to-day challenges of the position. The opportunities for discussion and reflection are especially valued in this form of training (Grissom & Harrington, 2010). This role can be seen as more of an apprenticeship or assistant principal role that takes place over an extended

period of time. Coaching differs from mentoring as it usually takes the form of an outsider observing and critiquing the work of the principal. There is a specific focus for the coaching and it takes place for a limited time (Grissom & Harrington, 2010).

Davis, et al. (2005) support the use of cohort groups as a means for professional development. Adult learning theory stresses the importance of adults taking an active role in their instructional improvement by working in a collaborative structure that fosters positive reinforcement and assistance (Danielson & McGreal, 2000). Cohorts help principals to build group and individual knowledge, think creatively, and to analyze problems from various perspectives (Davis, et al., 2005). The positive effects of cohort structured learning are: feelings of acceptance and group affiliation, social and emotional support, motivation, persistence, group learning, and mutual assistance (Davis, et al., 2005).

The Pittsburgh Principal Initiative Program (PPIP) is part of a comprehensive plan to improve the leadership in the Pittsburgh public schools. RAND Corporation evaluated the implementation of PPIP for four years and found that principals viewed the Directed Professional Growth (DPG) projects as the number one largest contributor to their professional development (Hamilton, Engberg, Steiner, Nelson & Yuan, 2010). DPG projects allow principals to choose an area from their evaluation rubric in which to work to improve their skills. The topics principals chose were related to improving student and teacher growth, relevant topics for principals implementing a new teacher evaluation system.

Barriers to participation in professional development. Professional development that school leaders receive is infrequent with a lack of cohesiveness to the

program (Donaldson, 2010; Wahlstrom, et al., 2010). In their research, Wahlstrom, et al. (2010) found that professional development had limited effects on principal efficacy and student achievement in school districts that had not developed and communicated clear goals for improvement. District sponsored professional development also had a negative effect on the efficacy of the principal when the district failed to acknowledge the unique needs of the school.

There are personal barriers to principals participating in professional development as well. One of the leading impediments is time (Barth, 1986; Sartain et al., 2011). Administrators feel the need to focus their training on day-to-day demands of administrative management competencies and less on curriculum and instructional practices. Hence, they have difficulty judging classroom instruction and tend to visit rarely (Fink & Resnick, 2001). Principals also do not want to reveal themselves as being flawed. The need to be a learner may reveal a lack of knowledge suggesting incompetence (Barth, 1986). Training for principals in the evaluation system should uncover and deal with the barriers that principals perceive (Painter, 2000a).

New programs for professional development opportunities for administrators are offering more on-the-job, experiential training and favor mentoring over book learning. They are long on practice and short on theory (Levine, 2005). They are more pragmatic, geared to specific knowledge and skills for principals at different stages of their careers. They can be in the form of study groups, seminars, reading and discussion groups, and presentations by expert practitioners or current thinkers. They can be Internet-based, streaming videos, or online discussions. Whatever form, professional development should include a strong component of coaching and feedback (Peterson, 2002). This

study may be helpful to district and intermediate units (IU) as they formulate their programs for supporting principals.

Summary and Implications of the Literature

While there have been substantial educational reforms since the publication of A Nation at Risk, there has not been a systematic, relevant, sustained professional development program for administrators. This literature review confirms the need for professional development for principals as a critical component to the success of the implementation of the new teacher effectiveness system.

Professional development can no longer be a sporadic workshop model. It needs be an ongoing system that addresses the specific needs of the principal (Peterson, 2002). Research on the knowledge, skills, and abilities principals need to be successful has not been well developed (Grissom & Harrington, 2010). More research-based evidence of the various means of delivering professional development will be useful for planners who are seeking to promote school improvement. Without this information, it will be difficult to construct a purposeful program for administrators and determine what format it should take (Grissom & Harrington, 2010).

In order to make teacher evaluation a beneficial process that leads to teacher growth, administrators must be knowledgeable of good teaching practices and have the skills necessary to be effective evaluators. The principal must also have the efficacy to believe that he/she can make a difference.

CHAPTER III

Methodology

The purpose of this chapter is to describe the research design, selection of the sample, instrumentation, and the data collection and analysis procedures in this study. The chapter is divided into five sections. Section one provides the statement of purpose for the study and research questions, section two describes the population and sample, section three describes the instrumentation, section four contains data collection procedures, and section five contains data analysis procedures.

Statement of Purpose

The primary purpose of this study was to determine the perceived professional development needs of Pennsylvania principals as they implemented the new educator effectiveness system. In doing so, this study examined the readiness of principals to implement the components of the Danielson Rubric in Domains 2 (The Classroom Environment) and 3 (Instruction). In addition, the study determined whether the professional development needs of principals were significantly related to personal demographics and organizational characteristics of the principals' schools. These include: the number of years of experience in the role of principal, the number of teachers evaluated annually, the school level (elementary, middle, high school), the percentage of poverty level students identified by the percentage of students receiving free and reduced lunch, the School Performance Profile percentage, and the model for teacher evaluation that the principal was using.

This study also examined the formats principals consider important in determining their participation in professional development activities. The study

examined the relationship between the delivery of services, such as workshops, online courses, university classes, study groups, or conferences and principals' personal demographics and organizational characteristics.

Research Questions

The following research questions guided this research study:

1. What is the confidence level of principals in their ability to evaluate teachers on the components of Domains (2) The Classroom Environment and (3) Instruction of the Danielson Framework?
2. What formats do principals prefer for participation in professional development activities related to implementing the educator effectiveness system?
3. What is the relationship between principals' personal demographics and their confidence regarding teacher evaluation?
    a. What is the relationship between principals' years of experience as a principal and their confidence regarding teacher evaluation?
    b. What is the relationship between the average number of teachers evaluated by the principals within a year and the principals' confidence regarding teacher evaluation?
4. What is the relationship between organizational characteristics and principals' confidence regarding teacher evaluation?
    a. What is the relationship between elementary, middle, and high school principals and their confidence regarding teacher evaluation?

b. What is the relationship between the percentage of free and reduced lunch students in a principal's school and principal's confidence regarding teacher evaluation?
c. What is the relationship between principals' School Performance Profile percentage and their confidence regarding teacher evaluation?

5. What is the relationship between personal and organizational characteristics and principals' preference regarding professional development formats related to implementing the educator effectiveness system?
    a. What is the relationship between principals' years of experience as a principal and their preference regarding professional development formats?
    b. What is the relationship between the average number of teachers evaluated by principals within a year and their preference regarding professional development formats?
    c. What is the relationship between elementary, middle, and high school principals and their preferences for professional development formats?
    d. What is the relationship between the percentage of free and reduced lunch students in principals' schools and principals' preferences for professional development?
    e. What is the relationship between principals' School Performance Profile percentages and their preferences for professional development?

Population and Sample

The target population for this study consisted of principals from the 500 public school districts in Pennsylvania excluding Philadelphia and charter schools. Given its overwhelming size and nature, the Philadelphia School District functions using a different organizational structure than that of other districts in the state. Similarly, charter schools do not function under the same regulations as public schools.

When considering the target sample size when estimating means, researchers look at statistical power. Gall, et al. (2007) indicate that several factors should be considered in statistical power analysis: sample size, level of significance, directionality, and effect size. Using the table provided by Kraemer and Theimann (1987), applying a two-tailed test (directionality), an alpha of .05 and a power of 80% (level of significance), and an effect size of .20, a sample size of 192 is sufficient.

A directory of principals was obtained through the Pennsylvania Department of Education Division of Data Quality. The population of principals in Pennsylvania is approximately 2565. Because doctoral student research typically receives approximately a 30% response rate to surveys (White, personal correspondence, August 2, 2014), a substantial oversampling of 1060 randomly selected principals was surveyed. After alphabetizing and numbering the names of principals, the sample was drawn using a random number generator. Emails of principals were gathered from district websites.

To enlist the support of the superintendents, an email was sent to the superintendents of identified participating school districts requesting that they encourage their principals to complete the survey in a timely manner. A copy of the results of the

survey will be sent to the superintendent of districts identified to participate in the research.

Another means to encourage participation was the inclusion of an incentive. Respondents who answered the survey completely were able to follow a link to a separate survey where they could leave their contact information in order to be included in a drawing for Amazon gift cards. Forty-five percent of the respondents participated in the incentive drawing.

Instrumentation

A closed-form, author designed instrument, The Principal Professional Development Scale (Appendix G), was developed using Danielson's Framework (Danielson, 2011a) and the research discussed in Chapter II (see Table 1). The questionnaire was divided into three parts. Following a set of directions, Part 1focused on Domains 2 and 3 of Danielson's Framework (Danielson, 2011a) and principals' efficacy in evaluating these domains. The two domains were broken down into four components each with specific elements for each of the eight components totaling twenty-eight elements. Part 2 asked principals to evaluate the formats they would pursue for professional development opportunities. Part 3 focused on the principals' personal demographic and organizational characteristics.

The Principal Professional Development Scale was developed based on the research discussed in the previous chapters. Table 1 provides the literature support for questionnaire items.

Table 1

Literature Support for The Principal Professional Development Scale

Question from Survey	Literature Support
Part 1	
Questions 1 through 28	Danielson, 2011a
Elements of the Danielson Framework	
Part 2	
Question 29 District or IU Workshops	Lashway, 2003; Hamilton et al., 2010
Question 31 Mentoring and coaching	Grissom and Harrington, 2010
Questions 32 through 34	Davis, et al., 2005; Grissom and Harrington, 2010; Levine, 2005
Question 35 Small study groups	Davis, et al., 2005
Part 3	
Question 39 Level of school	Leithwood and Jantzi, 2008b
Question 41 Years of experience.	Bryant, 2011; Kerrins and Cushing, 2000

The survey was piloted with 45 people familiar with the Framework for clarity, readability, and timing. See Appendix C for the introductory letter to participants and Appendix D for the survey to be completed. Participants in the pilot were practicing principals, assistant principals, administrative interns, staff developers, and curriculum coordinators. Feedback from pilot participants supported the refinement of the

questionnaire. This included the time required to complete the questionnaire, suggestions for revisions to each part of the questionnaire, and overall suggestions for revisions.

As a result of the pilot study, two questions were reworded and the number of possible choices was changed from four to six to increase the variability of the responses. It was also determined that the average time to complete the survey was approximately nine minutes.

Data Collection

Once the accessible population (also known as the sampling frame) was identified as described above, a cover letter along with the Principal Professional Development Questionnaire was electronically sent to superintendents and principals in the sample (Appendix E, F and G, respectively). The superintendents were asked to promote and support the research with their administrators. An electronic reminder was sent to follow-up with all participants two weeks following the initial request (Appendix H) asking that if they had not completed the survey that they please do so by the date specified.

In accordance with the Institutional Review Board standards, participation in this study was voluntary. Using the online survey software, Survey Monkey, participants were guaranteed anonymity. I had no way of tracking respondents to personally identifiable information through the Principal Professional Development Survey. A link to an incentive survey was embedded into the Principal Professional Development Survey. Respondents to the incentive survey could not be linked to the responses in the initial survey. Contact information for the incentive survey was completely voluntary. Participants were able to withdraw from the survey at any time. Of the 300 respondents,

279 completed the entire survey. All respondents were included in the data. Only aggregate data were reported in the final data analysis. The records from this study and any information collected through this research project were stored on a password-protected computer.

Data Analysis

Once all questionnaires were reviewed for clarity, quantitative statistical methods were employed to analyze the data. Descriptive statistics using measures of central tendency including the mean and measures of variability including standard deviation were used to measure the confidence level of principals in their ability to evaluate teachers on components 2 and 3 of Danielson's Framework. I summarized how likely or unlikely principals were to pursue specific formats when participating in professional development activities using means and standard deviations.

To examine the relationship between principals' personal demographics and principals' organizational characteristics (independent variables) and confidence levels and professional development formats (dependent variables), appropriate inferential analyses were used. When more than one dependent variable was used, as in the case of the component scores regarding principals' confidence level in teacher evaluation, a multivariate regression of each component score was examined. Comparing school level, a categorical variable, as the predictor (independent) variable and multiple outcome (dependent) variable scores from the confidence ratings, a multivariate analysis of variance, MANOVA, was used. Table 2 aligns the research questions with the source of data and the statistical analysis used to study the data.

Table 2

Research Questions and Methods of Statistical Analysis

Research Question	Data Source	Analytical Method
1. How confident are principals in their ability to evaluate teachers on Components 2 and 3 of the Danielson Framework?	Items 1 through 28 on the Principal Professional Development Scale	Descriptive statistics including means and standard deviations of each of the 8 components
2. How likely are principals to participate in various professional development formats?	Items 29 through 37 on the Principal Professional Development Scale	Descriptive statistics including means and standard deviations for items 36 through 44
3. What is the relationship between principals' personal characteristics and their confidence regarding teacher evaluation?	Items 41, 43, 1 through 28 on the Principal Professional Development Scale	Analyses are broken out by sub question below.
3a. What is the relationship between principals' years of experience as a principal and their confidence regarding teacher evaluation?	Items 41, 1 through 28 on the Principal Professional Development Scale	Multivariate regression of component scores on years of experience of the principal
3b. What is the relationship between the average amount of teachers evaluated by the principals in the past 5 years and the principals' confidence regarding teacher evaluation?	Items 43, 1 through 28 on the Principal Professional Development Scale	Multivariate regression of component scores on the average number of teachers evaluated
4. What is the relationship between principals' organizational characteristics and principals' confidence	Items 38 - 40 and 1 through 28 on the Principal Professional Development Scale	Analyses are broken out by sub question below.

regarding teacher evaluation?		
4a. What is the relationship between principals' school level (elementary, middle, and high school) and their confidence regarding teacher evaluation?	Items 38, 1 through 28 on the Principal Professional Development Scale	MANOVA with the predictor (independent) variable as the school level and the outcome (dependent) variable the scores on the component scores of the confidence ratings
4b. What is the relationship between the percentages of free and reduced lunch students in a principals' schools and the principals' confidence regarding teacher evaluation?	Items 39, 1 through 28 on the Principal Professional Development Scale	Multivariate regression of the component confidence scores on the percentage of students on free and reduced lunch
4c. What is the relationship between principals' School Performance Profile percentage and their confidence regarding teacher evaluation?	Items 40, 1 through 28 on the Principal Professional Development Scale	Multivariate regression of the component confidence scores on principals' School Performance Profiles
5. What is the relationship between principals' personal and organizational characteristics and principals' likelihood to pursue various professional development formats?	Items 29 through 41, 43, on the Principals Professional Development Scale	Analyses are broken out by sub question below.
5a. What is the relationship between principals' years of experience as a principal and their likelihood to pursue various professional	Items 41, 29 through 37 on the Principals Professional Development Scale	Separate regressions of professional development likelihood scores on the years of experience

development formats?		
5b. What is the relationship between the average amount of teachers evaluated by principals in the past 5 years and their likelihood to pursue various professional development formats?	Items 43, 29 through 37 on the Principals Professional Development Scale	Separate regressions of professional development likelihood scores on the average number of teachers evaluated
5c. What is the relationship between principals' school level (elementary, middle, and high school) and their likelihood to pursue various professional development formats?	Item 38, 29 through 37 on the Principal Professional Development Scale	Separate ANOVAs with school level as the predictor (independent) variable and likelihood to pursue various professional development formats as the outcome (dependent) variable
5d. What is the relationship between the percentage of free and reduced lunch students in principals' schools and principals' likelihood to pursue various professional development?	Item 39, 29 through 37 on the Principal Professional Development Scale	Separate regressions with the percent of students receiving free and reduced lunch as the predictor (independent) variable and likelihood to pursue various professional development formats as the outcome (dependent) variable
5e. What is the relationship between principals' School Performance Profile percentages and their likelihood to pursue various professional development formats?	Item 40, 29 through 37 on the Principal Professional Development Scale	Separate regressions with the School Performance Profile percentage as the predictor (independent) variable and likelihood to pursue various professional development formats as the outcome (dependent) variable

CHAPTER IV

Results

The purpose of this chapter is to report the findings of the study. First, descriptive data of the sample will be provided. The findings are then presented and analyzed to address each of the research questions.

In all, I sent the survey to 1060 principals from Pennsylvania. I also contacted superintendents to request their support for the research. Twenty-eight principals' e-mails were returned several times as undeliverable and 13 districts declined to participate, which excluded another 53 principals leaving a sampling frame of 1007. A total of 300 responses were gathered for a response rate of 31%.

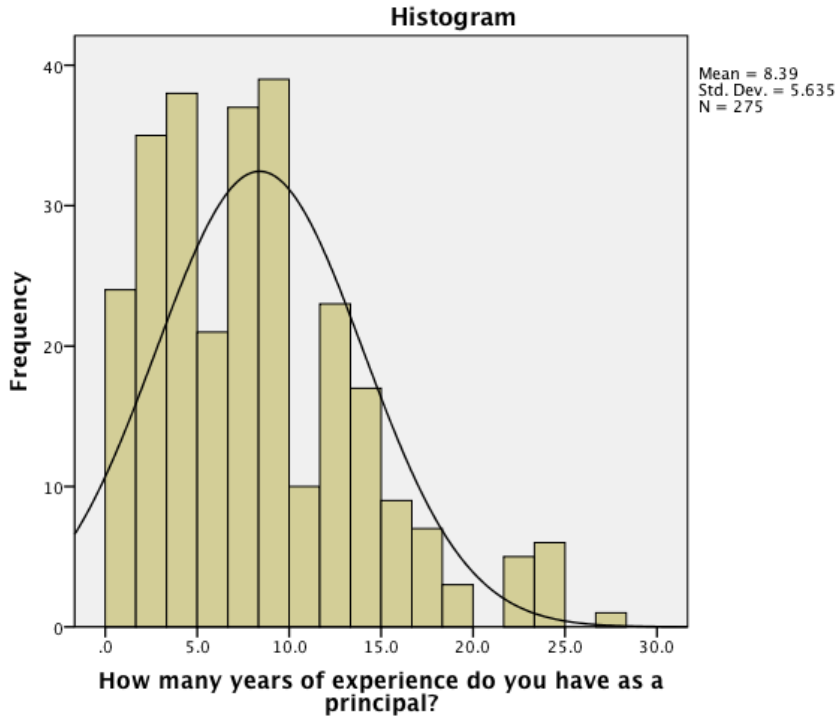
Descriptive Statistics for Personal and Organizational Data

Two hundred and seventy-five principals responded to the question regarding the number of years of experience the respondents had as a principal. The mean for the number of years respondents have been principals was 8.39. Because the median number of years as a principal was 8.00 years, about 50 % of the respondents fall between 0 and 8 years experience, and about 50 % of the respondents fall between 8 and 28 years experience. In the most recent statistics provided from the Pennsylvania Department of Education (2012 – 2013), the average number of years principals have served in their positions is approximately 18.7. This is consistent with the two previous years as well. (PDE, 2015c) This would indicate that the results from this study, 8.39 as the average number of years principals have served in their positions, is from a population with much less experience than the overall state average of 18.7. Therefore, the findings from this

study cannot be generalizable to the population in relation to the number of years of experience of the principal. Figure 1 shows the distribution of "years of experience."

Figure 1

Histogram of Data Showing the Number of Years as a Principal



Two hundred and seventy-two principals responded to the question regarding the average number of teachers they evaluate each year. The number of evaluations that principals conducted ranged from 0 to 93 with a mean of 34. Since the median was 30, about 50% of the respondents evaluated 30 or less teachers a year and about 50% evaluated more than 30 teachers a year. Of the twenty respondents who evaluated 60 or more teachers, 17 were secondary principals (12 high school, and 5 middle school). All but two respondents stated that they used the Danielson Framework to evaluate their teachers. Figure 2 shows the distribution of the "average number of teachers evaluated."

Figure 2

Histogram of Data Showing the Average Number of Teachers Evaluated

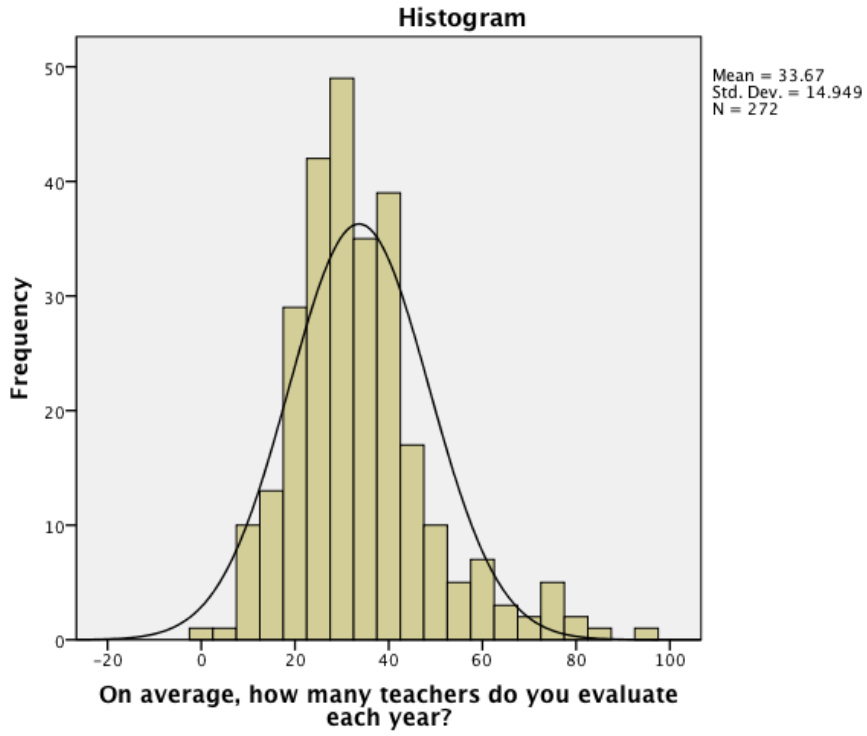


Table 3 presents the descriptive statistical data for principals' "school level."

Table 3

What level is your school?

		Frequency	Valid Percent
Valid	Elementary	172	62.1
	Middle	45	16.2
	High	60	21.7
	Total	277	100.0
Missing	System	23	
Total		300	

State-wide school level for elementary principals is 66% and secondary principals is 34%. (PDE, 2015c) This would indicate that this study with a 62% return rate from elementary principals and a 38% return rate from secondary principals is closely aligned to the general population in relation to school level.

In the most recent statistics provided by Pennsylvania Department of Education, the state-wide percentage of students on free and reduced lunch is 48.11% (PDE, 2015d). For the current study, on average, 38.69% of students were on free and reduced lunch. This would indicate that the schools in the current study do not perfectly represent the population of schools at large regarding socioeconomic status (SES). More specifically, the families in the schools of the current study are somewhat more affluent than the families in the general population in relation to poverty level. Thus, the findings from this study may not generalize to the poorest schools and districts in the state.

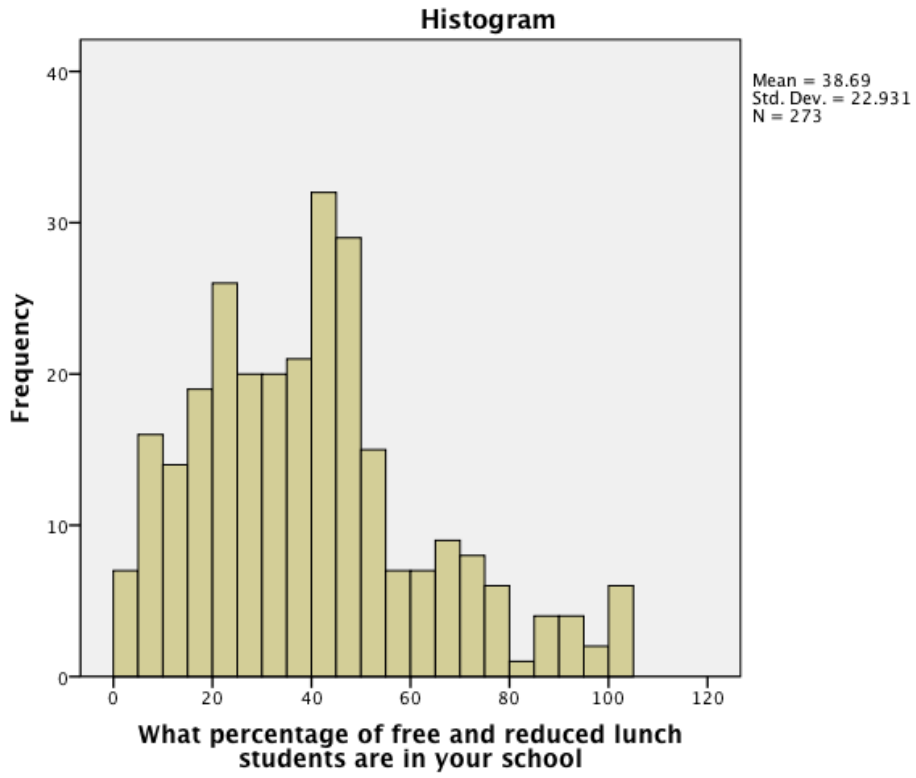
Part of the explanation for the SES difference between the sample and the population is that the school district of Philadelphia was not included in the sampling frame. Although I was unable to obtain the percentage of free and reduced lunch for the state excluding Philadelphia, it should be noted that Philadelphia is a large district serving many low SES communities with one hundred percent of their students eligible for free lunch (School District of Philadelphia, 2015). Without Philadelphia in the percentage of free and reduced lunch reported by the state, the sample may, in fact, be representative of the accessible population (Roberts, personal communication, July 19, 2015).

Figure 3 shows the percentage of free and reduced lunch for the schools represented by the respondents. The percentages range from 1 to 100 with the bar for 0-5% percentage of free and reduced lunch placed to the right of 0 and the bar for 95-

100% percentage of free and reduced lunch placed to the right of 100.

Figure 3

Histogram of Data Showing Percentage of Free and Reduced Lunch

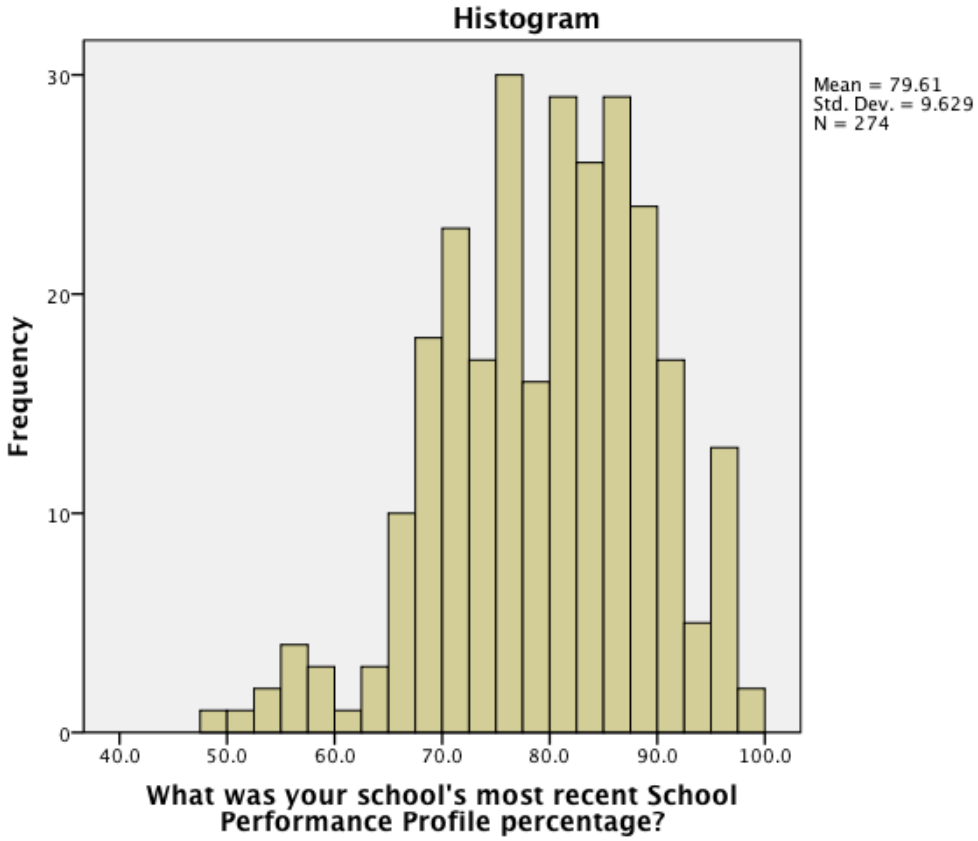


Two hundred and seventy four principals responded to the question regarding their School Performance Profile. The minimum percentage was 49.1% and the maximum was 99% with a mean of 79.1%. A median of 80.5% indicated that about 50 % of the respondents scored above 80.5% on the School Performance Profile and about 50% scored below 80.5%. Similar to the responses for the free and reduced lunch percentages, these percentages would indicate that most respondents were not from schools at the lower socio-economic level where the School Performance Profiles fall in the 60% range. (PDE, 2015e)

There is no state-wide average of School Performance Profile percentages. Not all schools have the same number of points available to them and different school levels have different weighting factors applied to different School Performance Profile elements. Therefore, it cannot be determined how the School Performance Profile relates to the general population of the state.

Figure 4

Histogram of Data Showing the School's Most Recent School Performance Profile Percentage



Findings

The research questions and results for this study are as follows:

Research Question 1: What is the confidence level of principals in their ability to evaluate teachers on the components of Domain 2, The Classroom Environment and Domain 3, Instruction of the Danielson Framework?

Likert scale responses on the survey were attributed numeric scores as follows: 6 = Strongly Agree, 5 = Agree, 4 = Slightly Agree, 3 = Slightly Disagree, 2 = Disagree, and 1 = Strongly Disagree. The highest mean scores indicate the elements principals feel most confident in evaluating. The lowest mean scores are the ones principals feel the least confident in evaluating. Table 8 provides the results of the survey.

Table 4

Means, Standard Deviations, and Sample Sizes for Principals' Confidence Levels Regarding Danielson's Domains 2 and 3

	n	*M*	*SD*
behavioral expectations.	289	5.19	.89
the communication of directions and procedures.	287	5.19	.82
the monitoring of student behavior.	287	5.18	.81
student participation	286	5.17	.82
the response to student misbehavior.	287	5.15	.95
the engagement of students using instructional materials and resources.	289	5.10	.85
teacher interactions with students, including both words and actions.	280	5.09	.84
student interactions with other students including both words and actions.	300	5.06	.86
the engagement of students in learning through activities and assignments.	289	5.05	.82

	n	*M*	*SD*
the management of transitions.	286	5.05	.88
the quality of questions and prompts.	287	5.03	.82
the structure and pacing of the lesson.	286	5.01	.87
teacher and student discussion techniques.	289	5.01	.82
the management of instructional groups.	289	5.00	.89
the management of materials and supplies.	287	4.99	.91
the communication of expectations for learning.	280	4.97	.85
the culture of high expectations for learning and achievement.	300	4.96	.89
the communication of content explanations.	289	4.94	.85
the use of oral and written language with students.	287	4.90	.83
assessment feedback to students.	300	4.88	.86
the lesson adjustment when needed.	300	4.86	.86
the assessment criteria.	289	4.83	.94
monitoring of student learning through assessment.	280	4.81	.84
the grouping of students.	286	4.77	.96
the teacher's ability to convey the importance of the content and of learning.	280	4.72	.86
the performance of non-instructional duties.	286	4.46	1.16
students' pride in their work.	286	4.30	1.03
the students' self-assessment and monitoring of progress.	280	4.26	1.03
Valid N (listwise)	280		

Overall, principals reported that they are fairly confident in evaluating teachers using Danielson's Framework Domains 2 and 3 with the mean of their confidence levels between "slightly agree" and just above "agree". Some of the highest confidence levels were in student behaviors: setting behavioral expectations, monitoring behaviors, and teachers' response to behaviors. All of these elements had mean scores that fell within the responses of "agree" and "strongly agree" and all are under component 2d: Managing Student Behaviors. Another component of high confidence was 2a: Creating an Environment of Respect and Rapport. The two elements in this component had mean scores that fell above the "agree" level.

The component principals had the least confidence in was 3d: Using Assessment in Instruction. All five elements fell within the ten least confident areas for principals. The element of least confidence was student self-assessment and monitoring of progress. Forty-six percent of principals agreed that they were confident evaluating this element, while fifty-four percent either slightly agreed or disagreed that they were confident assessing this element.

Another component principals were least confident in evaluating was 2b: Establishing a Culture of Learning. While confidence in evaluating the culture of high expectations for learning and achievement was higher than establishing the importance of the content and learning and student pride in work, all three elements were scored between slightly agree and agree. Nineteen percent of principals did not feel confident in evaluating a student's pride in his/her work.

Finally, the performance of non-instructional duties was in the bottom three elements in confidence and also had the highest standard deviation. Twenty percent of

principals were not confident in evaluating non-instructional duties. Principals had the least confidence in evaluating this element.

Research Question 2 - What formats do principals prefer for participation in professional development activities related to implementing the educator effectiveness system?

Likert scale responses on the survey were attributed numeric scores as follows: 4 = "Very Likely," 3 = "Likely," 2 = "Unlikely," and 1 = "Very Unlikely." The highest mean scores indicate the formats of professional development principals would most likely pursue. The lowest mean scores are the ones principals would be least likely to pursue. Table 5 provides the results of the survey.

Table 5

Means, Standard Deviations, and Sample Sizes for Principals' Preferences Regarding Professional Development Formats

	n	*M*	*SD*
District or IU sponsored workshops	279	3.36	.65
Mentoring and/or coaching sessions	279	2.90	.67
Small study groups	279	2.80	.69
State or National Conferences	279	2.75	.82
Independently sponsored self-paced online courses	279	2.51	.84
University course online	279	2.43	.86
University course work hybrid: on campus and partially online	279	2.42	.83
University course work on campus	279	2.39	.81
Valid N (listwise)	279		

District or IU sponsored workshops were the most likely format of professional development that principals would pursue with a mean of 3.36, which falls between the

likely and very likely range. The top three formats of professional development that principals would likely pursue are district or IU workshops, mentoring and/or coaching sessions, and small study groups.

The least preferred format for professional development was university course work on campus with a mean of 2.39. University course work online and a hybrid of online and on campus course work were also not preferred by principals with a mean of 2.43 and 2.42 respectively.

To determine if there was a significant difference between the top three formats for professional development and the bottom three formats, a pair-wise post hoc test was performed. First, the top three and bottom three formats were numbered in order of preference as seen in Table 6.

Table 6

Means, Standard Deviations, and Sample Sizes for Principals' Top 3 and Lowest 3 Preferences Regarding Professional Development Formats

Ordinal Preference Number		*n*	*M*	*SD*
1	District or IU sponsored workshops	279	3.36	.65
2	Mentoring and/or coaching sessions	279	2.90	.67
3	Small study groups[a]	279	2.80	.69
4	University course work online[b]	279	2.43	.86
5	University course work hybrid: on campus and partially online	279	2.42	.83
6	University course work on campus	279	2.39	.81

[a]The top 3 preferences appear above the horizontal line in the center of the table. [b]The lowest 3 preferences appear below the horizontal line in the center of the table.

Comparisons were then made between formats 1 and 2, 2 and 3, 3 and 4, 4 and 5, and 5 and 6. Because we increase the probability of finding a significant relationship by

chance alone each time we conduct a test, the alpha criterion needed to be adjusted using the Bonferroni adjustment. Dividing the alpha by the number of comparisons being conducted makes this adjustment. Therefore, dividing .05 by 5 gives an alpha of .01. A significant test must have an alpha of .01 or lower. Table 7 summarizes the results of these tests.

Table 7

Paired Samples Tests: Professional Development Formats

		Paired Differences		t	df	Sig. 2-tailed
		Mean	SD			
Pair 1	District or IU sponsored workshops - Mentoring and/or coaching sessions	.46	.77	10.01	278.00	.0005
Pair 2	Mentoring and/or coaching sessions - Small study groups	.10	.81	2.08	278.00	.04
Pair 3	Small study groups - University course online	.37	1.09	5.67	278.00	.0005
Pair 4	University course online - University course work hybrid: on campus and partially online	.01	.77	.31	278.00	.757
Pair 5	University course work hybrid: on campus and partially online - University course work on campus	.03	.70	.69	278.00	.494

In the first test comparing 1, district or IU sponsored workshops and 2, mentoring and or coaching sessions, there was clearly a significant difference (.0005) between these formats. In the second test, a significance of .04 was not significant between 2, mentoring and or coaching sessions, and 3, small study groups.

The third comparison of 3 and 4 presented a significant difference of .0005, indicating that principals have a stronger preference for small study groups over

university course work online.

Finally, the last two comparisons of 4 and 5, and 5, and 6, showed no significance with a Sig. of .757 and .494 respectively. These tests indicate that there were no significant differences between university course work whether it is online, on campus, or a hybrid of the two.

We can conclude that the preference for district or IU workshops was significantly stronger than for mentoring and/or coaching sessions, and there was no significant preference between mentoring and/or coaching sessions and small study groups. There was significant preference for small study groups over online university coursework, and there was no significant difference between any of the university professional development formats whether they were online, on campus, or a hybrid of online and on campus.

We can also conclude that the top three formats, district or IU sponsored workshops, mentoring and/or coaching sessions and small study groups were significantly more preferred than any of the university formats for professional development.

When asked for other options for professional development formats, there were 26 relevant responses. Twenty-three principals suggested small group or self-study learning opportunities. Principal professional learning communities and online information feeds (Twitter, blogging, and on-line modules) were suggested. Three respondents suggested collaboration opportunities with other school districts. One respondent wanted professional development in ways to help struggling teachers and also suggestions on how to help teachers wishing to attain the distinguished level.

Research Question 3: What is the relationship between principals' personal demographics and their confidence regarding teacher evaluation?

a. What is the relationship between principals' years of experience as a principal and their confidence regarding teacher evaluation?
b. What is the relationship between the average number of teachers evaluated by the principals within a year and the principals' confidence regarding teacher evaluation?

In order to make the analyses of this information more manageable, the elements were organized into groups. For a teacher to earn a distinguished score in many of the components, students are required to be active participants in the learning process. The four elements that measured student learning and ownership for the process were identified as: student interaction with other students, including both words and actions, student pride in work, student participation, and student self-assessment and monitoring of progress.

A principal components (PC) analysis (a special application of factor analysis) was examined which indicated the Keyser-Meyer-Olkin statistic (KMO) was .75 and the component accounted for most of the variance within the elements (60.34%). The scree plot revealed that these 4 items comprised a unidimensional construct. Although a KMO of .75 is considered "middling" according to Norusis (1994), a follow-up test of the Cronbach alpha coefficient (used to test the reliability of the scale) was conducted and found to be .77. This is an acceptable level for Cronbach alpha; the criterion for an acceptable Cronbach alpha coefficient is a value greater than .70. Taken together, all of these analyses provided a justification for creating a new variable called "student

ownership for learning" by computing the mean of the four elements in this scale. This will be referred to as **New Group 1: Student Ownership for the Learning**.

Next, a principal components (PC) analysis was conducted for the remaining elements in Domain 2 (all Domain 2 items with the exception of the "Student Ownership for Learning elements," i.e., student interactions with other students, including both words and actions and student pride in work). The PC analysis revealed a KMO statistic of .92 (deemed as marvelous by Norusis, 1994) and the component accounted for most of the variance within the elements (58%). Furthermore, the scree plot revealed these 10 items comprised a unidimensional construct. Finally, the Cronbach alpha coefficient (.92) reinforced the conclusion that these elements comprise a unidimensional construct. This cluster of elements will be referred to as **Domain 2A: Teacher Impact on the Learning Environment.**

The remaining elements for Domain 3 were clustered into a PC analysis (excluding Group 1: Student Ownership for Learning Environment elements, i.e. student participation and student self-assessment and monitoring of progress.) The KMO of .94 showed that these elements belong together and produce a "marvelous" component solution (Norusis, 1994). Moreover, the two components (groupings) explain most (62%) of the variance within the elements.

A varimax rotation was conducted that did not reveal a logical structure. This was followed by an oblimin rotation which found a logical structure for the elements as follows:

Table 8

Pattern Matrix[a]

	Groupings	
	1	2
3A2 communication of directions and procedures.	.912	
3C4 structure and pacing of the lesson.	.794	
3C3 engagement of students using instructional materials and resources.	.791	
3B1 quality of questions and prompts.	.776	
3B2 teacher and student discussion techniques.	.772	
3A1 the communication of expectations for learning.	.765	
3A3 communication of content explanations.	.765	
3A4 use of oral and written language with students.	.701	
3C1 engagement of students in learning through activities and assignments.	.613	
3D2 monitoring of student learning through assessment.		.904
3D3 assessment feedback to students.		.827
3D1 assessment criteria		.697
3C2 the grouping of students.		.604
3D5 lesson adjustment when needed.		.594

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 5 iterations.

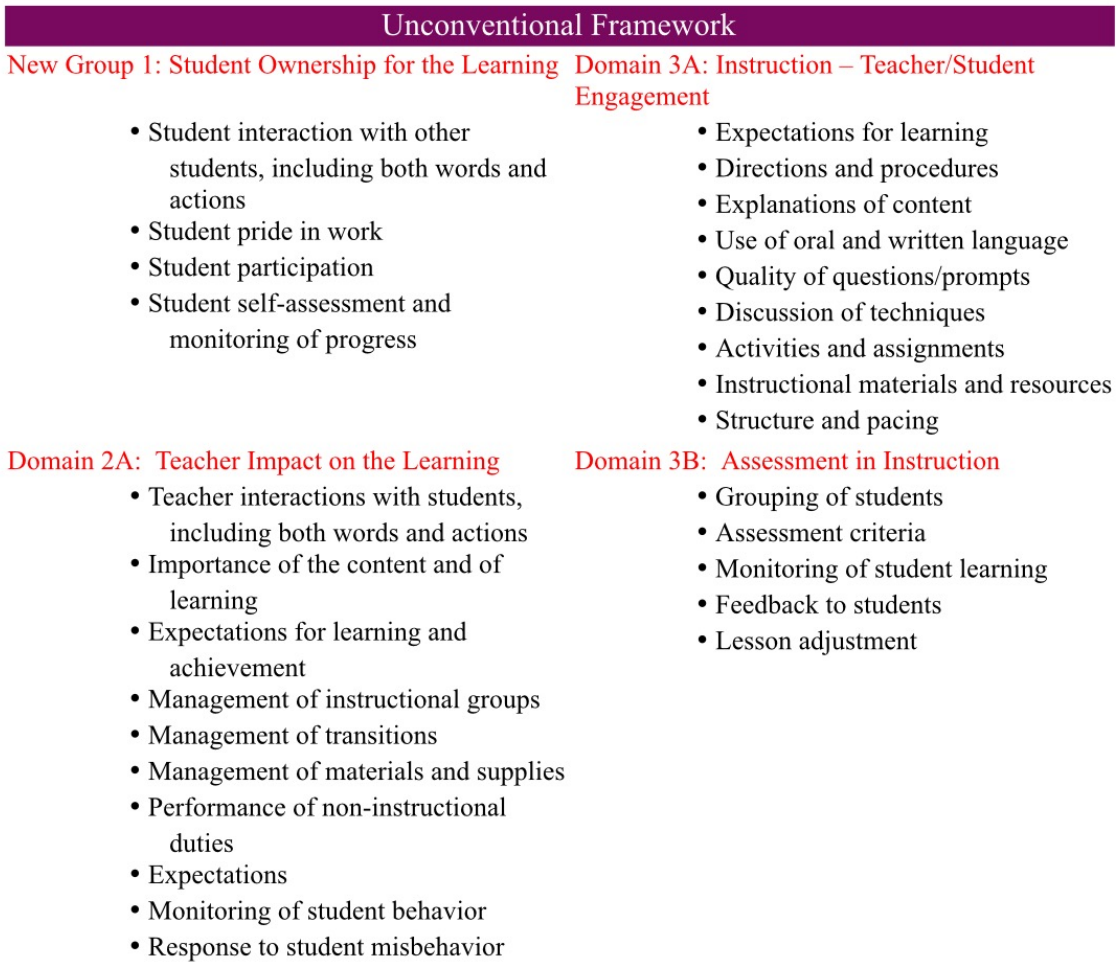
The result of the oblimin rotation showed two groups; group 1 was composed of all elements in 3a, 3b (except student participation), and most of the items in 3c (except grouping of students). A Cronbach alpha for these items was run and was found to be quite high (.92). These results provide a rationale for creating a new variable by averaging together the elements in this grouping. This new variable will be referred to as **Domain 3A: Instruction – Teacher/Student Engagement**.

The second grouping for Domain 3 was composed of all of the elements in 3d (with the exception of student self-assessment and monitoring of progress) and an element in 3c, grouping of students. The Cronbach alpha was computed for these 5 items and found it to be large (.85). Taken together, these findings provide a rationale for the decision to create a new variable by averaging these elements together. This new variable will be referred to as **Domain 3B: Assessment in Instruction.**

All of the elements in the original Framework were redistributed into unconventional, but logical groupings. From this point forward, the analysis of data was based on the new groupings. In order to better visualize the organization of these newly formed domains, the domains have been reformatted into Figure 5.

Figure 5

Unconventional, but Logical Grouping of Elements



After analyzing these elements in the new groupings, Table 9 presents the descriptive statistical data for the new variable groupings.

Table 9

Means, Standard Deviations, and Sample Sizes for Principals' Confidence Levels Regarding the New Variable Groups

		New Group 1 Student Ownership for Learning Environment	Domain 2A Teacher Impact on Learning Environment	Domain 3A Instruction Teacher/Student Engagement	Domain 3B Assessment in Instruction
N	Valid	300	300	289	300
	Missing	0	0	11	0
Mean		4.72	4.99	5.02	4.83
Median		4.75	5.00	5.00	5.00
Std. Deviation		.748	.703	.657	.713
Minimum		1.00	1.60	1.11	1.60
Maximum		6.00	6.00	6.00	6.00

Using these groupings of the elements, principals continued to show confidence in evaluating the elements in Domains 2 and 3 of the Framework. These results indicated that principals are most confident in evaluating the instructional factors of a lesson with a mean of 5.02, which corresponds to a qualitative response of "agree," (i.e. "I agree with the statement, 'I am confident.') Principals were least confident in the elements where the students respond to the instruction 4.72, which corresponds to a qualitative response between "4 = slightly agree" and "5 = agree" but was closer to "agree" (i.e., I agree/slightly agree with the statement, 'I am confident.').

A paired samples test was used to demonstrate the relationships between the domains in relation to their mean scores. It was found that all pairwise comparisons were significant, indicating that there is little overlap between the mean scores. It can then be concluded that domain 3A is greater than 2A, which is greater than 3B, which is greater than New Group 1.

Table 10

Paired Samples Tests: New Groupings

		Paired Differences	t	df	Sig. (2-tailed)
		Mean			
Pair 1	Student - Domain2A	-.27	-10.09	299	.0005
Pair 2	Student - Domain3A	-.31	-11.30	288	.0005
Pair 3	Student - Domain3B	-.11	-3.59	299	.0005
Pair 4	Domain2A - Domain3A	-.04	-2.19	288	.03
Pair 5	Domain2A - Domain3B	.16	5.26	299	.0005
Pair 6	Domain3A - Domain3B	.19	6.21	288	.0005

Using the new groupings I analyzed the data to address the specific characteristics.

a. What is the relationship between principals' years of experience as a principal and their confidence regarding teacher evaluation?

Step 1 – In checking the assumption of the test that the outcome variables are well correlated, a Bartlett's test was used showing a significant probability level ($p < .05$), which allows the conclusion that the assumption has been met.

Step 2 – Examining the multivariate effect using the Pillai's Trace (see Table 11). The "Sig." level was less than .05. This allows the researcher to examine the univariate tests to determine which of the 4 outcome variables were linked to "years as principal."

Table 11

Multivariate Test for the Link between Confidence Regarding Teacher Evaluation and Years as a Principal

Effect		Value	F	Sig.	Partial Eta Squared	Observed Power[c]
yearsprin	Pillai's Trace	.043	3.008	.019	.043	.796

For these results the effect size (Partial Eta Squared) was small. To put this

statistic in context, .02 is small .06 is medium, and .13 is large. Observed Power is the probability that I would find a significant effect in the data from my sample, should there be a real effect in the population studied. In these results, there was approximately an 80% chance of finding a significant effect in the sample, should there be a real effect in the population studied. However, there was also approximately a 20% chance of making a type 2 error of missing a significant effect when in fact there was one.

Step 3 – Analysis of Univariate Effects – Since the "Sig." level was less than .05, a relationship has been established between the variables. The Partial Eta Squared level showed the effect size that principals' years of experience have on their confidence regarding teacher evaluation.

Table 12 *Univariate Test for the Link between Confidence Regarding Teacher Evaluation and Years as a Principal*

Dependent Variable	Parameter	B	Std. Error	t	Sig.	Partial Eta Squared	Observed Power
Group 1 Student Ownership for Learning Environment	yearsprin	.019	.008	2.41	.017	.02	.671
Domain 2A Teacher Impact on Learning Environment	yearsprin	.022	.007	2.95	.003	.03	.837
Domain 3A Instruction Teacher/Student Engagement	yearsprin	.022	.007	3.14	.002	.03	.879
Domain 3B Assessment in Instruction	yearsprin	.011	.008	1.41	.159	.01	.291

There was a significant, positive link between "years as a principal" and three of the

four outcome variables (Group 1: Student Ownership for Learning Environment, Domain 2A: Teacher Impact on Learning Environment, and Domain 3A: Instruction Teacher/Student Engagement). For all of these results the effect size (Partial Eta Squared) was small. To put this statistic in context, .02 is small .06 is medium, and .13 is large. In other words, the more years a principal has been evaluating teachers, the more confident the principal is in evaluating the elements in Group 1, Domain 2A, and Domain 3A.

b. What is the relationship between the average number of teachers evaluated by the principals within a year and the principals' confidence regarding teacher evaluation?

Step 1 – In checking the assumption of the test that the outcome variables are well correlated, a Bartlett's test was used showing a significant probability level ($p < .05$), which allows the conclusion that the assumption has been met.

Step 2 – Examining the multivariate effect using the Pillai's Trace (see Table 13). The "Sig." level was greater than .05 indicating that the multivariate test was not significant. The analysis of this predictor was ended and there was no need to move to an analysis of the univariate effects.

Table 13

Multivariate test for the Link between Confidence Regarding Teacher Evaluation and the Number of Teachers Evaluated

Effect		Value	F	Sig.	Partial Eta Squared	Observed Powe
Teachers evaluated	Pillai's Trace	.028	1.922	.107	.028	.576

It was concluded that there was no correlation between the average number of teachers a principal evaluated in a year, and the principals' confidence level in applying the Framework to evaluate teachers.

Research Question 4: What is the relationship between organizational characteristics and principals' confidence regarding teacher evaluation?

a. What is the relationship between elementary, middle, and high school principals and their confidence regarding teacher evaluation?

Step 1 – In checking the assumption of the test that the outcome variables are well correlated, a Bartlett's test was used showing a significant probability level ($p < .05$), which allows the conclusion that the assumption has been met.

Step 2 – Examining the multivariate effect (see Table 14) it was noted that the "Sig." level was greater than .05 indicating that the multivariate test is not significant. The analysis of this predictor was ended and there was no need to move to an analysis of the univariate effects.

Table 14

Multivariate Test for the Link between Confidence Regarding Teacher Evaluation and School Level

Effect	Value	F	Sig.	Partial Eta Squared	Observed Power
School level	.03	.93	.487	.014	.44

The data indicate that there was no correlation between the level where principals practice, elementary, middle, or high school, and the principals' confidence level in applying the Framework to evaluate teachers.

b. What is the relationship between the percentage of free and reduced lunch students in a principals' school and principals' confidence regarding teacher evaluation?

Step 1 – In checking the assumption of the test that the outcome variables are well correlated, a Bartlett's test was used showing a significant probability level ($p < .05$), which allows the conclusion that the assumption has been met.

Step 2 – I examined the multivariate effect using the Pillai's Trace see Table 15. The "Sig." level was greater than .05 indicating that the multivariate test was not significant. The analysis of this predictor was ended and there was no need to move to an analysis of the univariate effects.

Table 15

Multivariate Test for the Link between Confidence Regarding Teacher Evaluation and the Percentage of Free and Reduced Lunch

Effect		Value	F	Sig.	Partial Eta Squared	Observed Power
Free and Reduced Lunch	Pillai's Trace	.03	2.33	.060	.034	.67

It was concluded that there was no correlation between the percentage of free and reduced lunch students in principals' schools, and the principals' confidence level in applying the Framework to evaluate teachers.

c. What is the relationship between principals' School Performance Profile percentage and their confidence regarding teacher evaluation?

Step 1 – In checking the assumption of the test that the outcome variables are well correlated, a Bartlett's test was used showing a significant probability level ($p < .05$), which allows the conclusion that the assumption has been met.

Step 2 – I examined the multivariate effect using the Pillai's Trace (see Table 16). The "Sig." level was less than .05. This allowed me to examine the univariate tests to determine which of the 4 outcome variables were linked to "School Performance Profile percentages."

Table 16

Multivariate Test for the Link between Confidence Regarding Teacher Evaluation and School Performance Profile

Effect		Value	F	Sig.	Partial Eta Squared	Observed Power
Performance Profile	Pillai's Trace	.05	3.79	.005	.053	.89

Step 3 – Analysis of the Univariate Effects – A "Sig." level of less than .05 showed a link between the "School Performance Profile" and the principal's confidence regarding teacher evaluation.

Table 17

Univariate Test for the Link between Confidence Regarding Teacher Evaluation and the School Performance Profile

Outcome Variable	Parameter	B	Std. Error	t	Sig.	Partial Eta Squared	Observed Power
Group 1 Student Ownership for Learning Environment	Performance Profile	.008	.004	1.96	.06	.014	.496
Domain 2A Teacher Impact on Learning Environment	Performance Profile	.011	.004	2.70	.007	.026	.768
Domain 3A Instruction Teacher/Student Engagement	Performance Profile	.014	.004	3.47	.001	.042	.933
Domain 3B Assessment in Instruction	Performance Profile	.012	.004	2.87	.004	.029	.815

There was a significant, positive link between "School Performance Profile" and three of the four outcome variables (confidence in evaluating Domain 2A, Domain 3A, and Domain 3B). The effect sizes for Domain 2A and Domain 3B were small. The effect size for Domain3A was small to medium. In other words, for these three outcome variables, principals with higher performance profile scores felt more confident in their evaluations of teachers.

Research Question 5: What is the relationship between personal and organizational characteristics and principals' preference regarding professional development formats related to implementing the educator effectiveness system?

a. What is the relationship between principals' years of experience as a principal and their preference regarding professional development formats?

Table 18 presents the relationship between the "years as a principal" and the principals' preference for professional development formats.

Table 18

Results of Inferential Tests of the Links between Years as Principal and Preference for Professional Development Formats

Professional Development Format	Standardized *B*	*t*	*p*
District or IU sponsored workshops	.01	.09	ns
Independently sponsored self-paced online courses	.03	.48	ns
Mentoring and/or coaching sessions	-.00	-.03	ns
University course work on campus	-.03	-.48	ns
University course work hybrid: on campus and partially online	-.07	-1.21	ns
University course online	-.09	-1.57	ns
Small study groups	.09	1.49	ns
State or National Conferences	.05	.78	ns

Table 18 indicates that "years as principal" was not linked to preferences for any of the professional development formats. This is evidenced in the column with the heading *p,* that stands for "probability level." The symbol "ns" in each cell shows that the probability of a correlation was "not significant" in every case.

b. What is the relationship between the average number of teachers evaluated by principals within a year and their preference regarding professional development formats?

Table 19 presents the relationship between the "number of evaluations" and the principals' preference for professional development formats.

Table 19

Results of Inferential Tests of the Links between Number of Evaluations and Preference for Professional Development Formats

Professional Development Format	Standardized *B*	*t*	*p*
District or IU sponsored workshops	.01	.20	ns
Independently sponsored self-paced online courses	.02	.37	ns
Mentoring and/or coaching sessions	-.03	-.56	ns
University course work on campus	.01	.19	ns
University course work hybrid: on campus and partially online	-.02	-.33	ns
University course online	-.00	-.01	ns
Small study groups	.00	.08	ns
State or National Conferences	.04	.74	ns

Table 19 indicates that the "average number of teachers evaluated" was not linked to preferences for any of the professional development activities. This is evidenced in the column with the heading *"p."* This statistic stands for "probability" that the correlation is significant. In all cases, the correlation between "number of evaluations" and "preference for a given professional development activity" was "not significant (ns)."

c. What is the relationship between elementary, middle, and high school principals and their preferences for professional development formats?

Table 20 represents the relationship between principals' "school level" and their preference for specific professional development formats. A *p* score of less than .05 is considered significant.

Table 20

Results of Inferential Tests of the Links between School Level and Preference for Professional Development Formats

Professional Development Format	*F*	*p*
District or IU sponsored workshops	1.51	ns
Independently sponsored self-paced online courses	1.29	ns
Mentoring and/or coaching sessions	3.03	.05
University course work on campus	0.22	ns
University course work hybrid: on campus and partially online	0.89	ns
University course online	0.31	ns
Small study groups	3.48	.04
State or National Conferences	2.68	ns

Table 20 indicates that there was a significant effect for "mentoring and/or coaching sessions" and for "small study groups." None of the pairwise post hoc tests (see Table 21) for "mentoring and/or coaching session," however, reached the level of significance. With regard to "small study groups," elementary principals had a stronger preference than did high school principals.

Table 21

Means, Standard Deviations, and Sample Sizes for School Level and Principals' Preferences Regarding Professional Development in Small Study Groups

	N	Mean	Std. Deviation
Elementary	172	2.84	.69
Middle	45	2.91	.67
High	60	2.60	.69
Total	277	2.80	.69

Both elementary and high school principals' scores are between "likely (3)" and "unlikely (2)," but the elementary principals' score was closer to "likely." The mean for

middle school principals was also close to "likely," but was not found to be significantly different than the high school principals' mean. This could be due to a type 2 error caused by the smaller sample size, and subsequent lower power of the test.

d. What is the relationship between the percentage of free and reduced lunch students in principals' schools and principals' preferences for professional development?

Table 22 represents the inferential statistics that show the correlations between "free and reduced lunch" and principals' preference for specific professional development formats. A *p* score of less than .05 is considered significant.

Table 22
Results of Inferential Tests of the Links between Free and Reduced Lunch (%) and Preference for Professional Development Formats

Professional Development Format	Standardized *B*	*t*	*p*
District or IU sponsored workshops	.06	1.06	ns
Independently sponsored self-paced online courses	.13	2.23	.03
Mentoring and/or coaching sessions	-.10	-1.65	ns
University course work on campus	.06	1.06	ns
University course work hybrid: on campus and partially online	-.02	-.36	ns
University course online	.01	.23	ns
Small study groups	-.11	-1.90	ns
State or National Conferences	-.01	-.12	ns

The data in Table 22 indicates that there was a significant, positive link between "free and reduced lunch" and preference for "independently sponsored self-paced online courses." More specifically, in schools with more low-socioeconomic students, principals were more likely to desire "independently sponsored self-paced online courses."

e. What is the relationship between principals' School Performance Profile percentages and their preferences for professional development?

Table 23 represents the inferential statistics that show the correlations between "School Performance Profile" and principals' preference for specific professional development formats. A *p* score of less than .05 is considered significant.

Table 23

Results of Inferential Tests of the Links between School Performance Profile and Preference for Professional Development Formats

Professional Development Format	Standardized *B*	*t*	*p*
District or IU sponsored workshops	-.00	-.02	ns
Independently sponsored self-paced online courses	-.06	-1.06	ns
Mentoring and/or coaching sessions	.04	.61	ns
University course work on campus	.04	.67	ns
University course work hybrid: on campus and partially online	.12	1.94	ns
University course online	.02	.38	ns
Small study groups	.13	2.17	.04
State or National Conferences	.12	1.94	ns

According to Table 23, there was a significant, positive link between "school performance profile" and preference for small study groups. More specifically, in schools with higher performance profiles, principals showed greater preference for participation in small study group activities.

Chapter V

Discussion and Recommendations

Summary

The findings of this study suggest that principals who participated in this study were confident in assessing the elements in Domains 2 and 3 of the Danielson Framework. Principals have the most confidence in evaluating Component 2d: Managing Student Behavior and 2a: Creating an Environment of Respect and Rapport. The component principals' have the least confidence evaluating is 3d: Using Assessment in Instruction. When the elements were examined in unconventional, but logical groupings, it was found that assessment in instruction continued to be the grouping that principals had less confidence in evaluating. Interestingly, elements associated with student ownership of the learning were the elements that principals had the least confidence in evaluating. The elements in this grouping are: student interactions with other students, including both words and actions, students' pride in their work, students' self-assessment and monitoring of progress, and student participation. This study extended the literature on teacher evaluation by recognizing that principals are confident in evaluating teachers using the Framework and identifying that evaluating teachers based on the actions of their students as indicated through classroom observations may be an area to examine in more depth.

This study also extended the literature base by identifying the preferred professional development formats in which principals would be most willing to participate. It was found that district and I.U. sponsored workshops were the preferred format for professional development. University course work, whether on campus,

online, or a hybrid of the two, was the least favorable means of professional development for principals.

The implications of this study were two-fold. First, the findings suggest that those providing professional development for principals would do well to examine how principals are able to collect evidence to support the students' learning and participation in the learning process. Also, professional developers, including school districts, intermediate units, and universities, should offer professional development that is relevant to the demographic population through workshops, mentoring/coaching sessions, or small study groups.

Discussion

Confidence levels. Researchers found that the training of principals in the use of Danielson's Framework was a key element in the success of implementing this standards-based evaluation system (Danielson, 2007; Kimball et al., 2004; Marshall, 2005; Sartain, et al., 2011; Sawyer, 2001). Findings from this study revealed that generally principals are confident evaluating teachers using Domains 2 and 3 of Danielson's Framework. A pattern that emerged from the research was that principals were most confident in evaluating student behaviors.

Component 2d: Managing Student Behaviors had the highest mean score indicating that principals are the most confident in evaluating the success of these elements. While many student behaviors are overt, determining what students are thinking during a lesson is not as easily quantified.

Student ownership of the learning environment determines the difference between a proficient teacher and a distinguished teacher (Danielson, 2007). In my experience as a

principal, what is more challenging than evaluating the teacher's instruction is evaluating the student's learning and ownership for the learning. I hypothesized that certain elements related to student ownership for the learning and student involvement in the learning process should be examined as a cluster. Taking four elements that specifically give students ownership for the learning, a new component, labeled Group 1, was formed. Principals indicated less confidence in evaluating Group 1: Student Ownership for the Learning Environment, than any of the other components.

Student self-assessment and a student's pride in his/her work were the two elements in Group 1 that principals had the least confidence in evaluating. This suggests that a teacher's evaluation may need to go beyond what is happening during a lesson. Principals may need to examine student assessment results (pre and post tests), student work samples, rubrics, and student self-reflections in order to assess whether the teacher is meeting the expectation for these elements. What could be gained by post conferencing with the student? The research of Gentilucci and Muto (2007) indicates that principals who check on the work of students have a more powerful influence on student achievement than principals who focus on managerial tasks. Post conferencing with students and analyzing their understanding of the teacher's instruction will bring the principal closer to the learning. This may result in higher student achievement and a more accurate teacher evaluation.

The results of another grouping of elements labeled, Domain 3A: Instruction and Teacher/Student Engagement, indicated that principals are confident in recognizing instructional best practices, such as: activities and assignments that intellectually engage students, the use of questioning that promotes higher level thinking, and communicating a

clear objective for the lesson. Being able to collect tangible evidence to support clinical feedback, enables the principal to keep personal judgment out of the evaluation (Danielson & McGreal, 2000; Ovando, 1994; Wiggins, 2012).

Research on the principals' years of experience and its effect on a principals' effectiveness is mixed. Leithwood and Jantzi (2008) found no significant effect from years of experience as a leader on a principal's efficacy. Bryant (2011) found that principals with six or fewer years as a principal were less effective with the teacher evaluation process. This current study found a correlation between principal efficacy and years of experience to three of the four domains studied. Unfortunately, the effect size is so small that it is difficult to lend meaning to the results. Principals with more experience may have a better understanding of the diverse instructional strategies due to the increased time spent in classrooms. In both Bryant's research and this study, the small effect sizes may be attributed to more than 50% of respondents having less than 10 years of experience as a principal.

Of interest in the findings was that principals with higher School Performance Profiles were more confident evaluating teachers in three of the components of the Framework. Since it has already been shown that principals in this study were more confident in assessing instructional practices, this research may add to the meta-analysis completed by Robinson, Lloyd, and Rowe (2008), where it was found that leaders in high performing schools worked directly with teachers to evaluate their teaching (Robinson, et al., 2008). As principals refine instructional practices with their teachers, student achievement scores should increase.

Professional Development. Principals rated district or Intermediate Unit (IU) sponsored workshops, mentoring and/or coaching sessions, and small study groups as the three most favorable formats for professional development. All of these formats are local and provide collegial, small group or one–on-one experiences. Local IUs offer principals opportunities to work with colleagues who share regional demographics, require less travel time, and provide access to local resources.

The formats for professional development principals found least favorable were university course work on campus, online, or a hybrid of both. These findings add to the research of Levine (2005) in which principals reported that leadership preparation programs were not relevant to their work. Tuition and travel may have been factors in principals' lack of interest in university course work. What also must be taken into consideration is that the purpose of university course work may not be to teach the specific skills of the professional development principals require. Principals may view university courses as a means to understand the theory and over-arching philosophy of education rather than a format for developing specific skills. It would be beneficial to delineate the difference between university course work and university sponsored professional development workshops.

Examining the organizational characteristics of principals, it was found that all three factors: school level, percentage of free and reduced lunch, and School Performance Profile percentage, showed a correlation to preferred professional development formats.

First, the school level, elementary, middle, and high, showed a relationship with the format of mentoring and/or coaching and the format of small study groups. Both of

these formats for professional development support the specific areas of need for principals. When professional development is not relevant to the needs of the principal, it will be of limited use (Wahlstrom, et al., 2010).

There also was a positive correlation between the schools with a higher percentage of low-socioeconomic students and principals' preference for independently sponsored self-paced online courses. It may be suggested that financial resources are a factor in principals' preferences. There are online courses and opportunities for professional development that are free or available at a very low cost. Another factor that may support the preference for independently sponsored self-paced online courses is time. Schools in low-socioeconomic areas where schools have a tendency to be underfunded result in the leadership taking on more responsibilities than their colleagues in schools where funding allows for the hiring of more support personnel.

Finally, principals in schools with higher School Performance Profiles (SPP) prefer to participate in small study groups. This form of professional development encourages collaborative problem solving and critical thinking to analyze methods to improve instruction (Davis, et al., 2005). In Cotton's (2003) review of 81 reports on principal behaviors that positively affect student achievement, collaboration was found to be one of the characteristics. This finding suggests that schools with a higher SPP may foster a culture of collaboration and collective responsibility.

Limitations

Three limitations surfaced as a result of the analysis of the data in this research. First, there was limited participation from school districts with a high percentage of students receiving free and reduced lunch from across the state. Also, the definition of

evaluation was not clear on the survey and may have caused a misrepresentation of the data. The third limitation was in the use of a self-report survey as a single source of data.

Although the number of responses to this survey was more than adequate to establish statistical significance and there was overwhelming support from superintendents, missing from the data was the participation of school districts with high percentages of students receiving free and reduced lunch. Philadelphia School District was not included due to its differing organizational structure. Allentown School District required an extensive application process for participation in a research study that did not fit within the timeframe of the survey. And, participation from other low socio-economic districts was limited based on evidence from the percentage of free and reduced lunch and SPP scores reported. Under representation of these low socio-economic schools limits the generalizability of the findings regarding principals' confidence levels, as well as their preferences for professional development activities. The findings generalize primarily to principals in schools with middle and upper class students.

Having piloted the Framework through the MET project, Pittsburgh reported in 2011 that principals did not feel confident implementing the Framework (Lane & Horner, 2011). More participation by the Pittsburgh School District may have added to the research of the MET project either by confirming that lack of confidence or showing an increase in confidence from the inception of the program.

Another limitation to this research was the interpretation by the respondents of the average number of evaluations performed each year. Principals complete a summative evaluation for all teachers each year, however summative evaluations only use the domains and not the elements that were used in this survey. I hypothesized that the more

principals evaluated teachers using the elements from the Framework, the more confidant principals would be collecting evidence from teacher observations. The question on the survey was poorly worded and did not distinguish between formative assessment using the elements of the Framework and summative assessment using just the domains.

Questions 3b and 5b found no correlation between the number of teachers evaluated annually and principals' confidence levels or professional development formats. In retrospect, if this question were to be asked again, it would be beneficial to ask principals how many teachers they formatively assess using the elements of the Framework. This number would most likely be substantially lower than the number of teachers summatively assessed using only the domains, especially in school districts in which a differentiated supervision model is used. With the data collected from this question, we must interpret the findings as there is no correlation between principals' confidence levels in evaluating teachers and the average number of teachers summatively evaluated using the domains.

Finally, a limitation of this survey was in the use of a self-report study. This type of survey is most effective when the respondents are knowledgeable about the subject of the survey and the questions are worded clearly (Gonyea, 2005). Social desirability bias, where a respondent presents oneself in a socially acceptable manner (Gall, et al., 2007), may have threatened the credibility of this self-report data. Principals may have fallen into this pattern if they believed their role required them to be proficient in evaluating the elements.

To reduce the effect of this bias, Nancarrow and Brace (2000) suggest that the survey be conducted using a computer rather than a face to face interview, that the

researcher appeals to the honesty of the respondent, and assures confidentiality and anonymity. All of these were included as part of this research. To encourage principals to give more thought to their responses, additional response options were included (Slightly Agree and Slightly Disagree) following the pilot study. An even number of responses (forced choice method) was purposely chosen to influence the respondents into either agreeing or disagreeing.

To further account for this bias, the triangulation of data sources would help to corroborate the results. Follow-up interviews with principals would have allowed for confirmation that principals indeed had knowledge of evidence used in evaluating the elements. Also, a survey of teachers' perceptions of principals' abilities to evaluate the evidence would have added to the validity of the data.

In addition to the limitation of social desirability bias, this research reported principals' perceptions of their abilities to evaluate teachers using the Framework rather than principals' actual competence in using the Framework. Seeing the results of actual observation feedback, whether it is a post conference write-up or witnessing a post conference discussion, would give data to support a principal's perception of confidence.

Recommendations for Practice

This study suggests the need to examine the process for gathering evidence and identifying the evidence itself for students' ownership for the learning. Specific emphasis should be on how students demonstrate pride in their work and how students assess and monitor their progress. The process may include conferencing with students and/or examining student work products. Ultimately, the goal would be to identify strategies that teachers can use to enhance these skills for students.

As school districts wrestle with the implementation of the Framework, they may want to examine the evidence used to evaluate Component 3d: Using Assessment in Instruction. Universities may want to give this more emphasis in their preparatory classes for administrators, so new principals will have a better understanding of what it should look like and sound like in the classroom. The study and implementation of formative assessment techniques, such as the use of white boards, exit slips, and electronic assessments, will not only help teachers assess the learning, but also give feedback to students as to their level of understanding of the lesson. In addition to formative assessment techniques, the ability to use the information gained from the assessments to formulate instruction will be a powerful tool for teachers.

Districts and intermediate units should be aware of their responsibilities in providing relevant professional development for principals. Principals are counting on these entities to keep them informed of current trends in education. Two characteristics of professional development were brought to light through this research: collaboration with colleagues and the desire for electronic, flexible opportunities to learn.

Collaboration with colleagues was a theme of the open-ended responses. The desire for professional learning communities reinforced the research of Danielson and McGreal (2000) who reported that collaboration promoted professional reflection while offering the perspective of colleagues.

Respondents also suggested non-traditional formats for professional development that were electronic-based and/or involved the use of the Internet. These included on-line modules, Personal Learning Network (PLN), Twitter, blogging, social media, and web sites, such as Edcamp. These relatively new formats for learning are convenient, time

efficient, and self-paced. With principals serving as instructional leaders along with managing a building, and in some cases two buildings, it is no surprise that time and convenience may influence the principals' choices for professional development. Barth (1986) and Sartain et al. (2011) confirm that time is one of the foremost barriers for principals' participation in professional development. Self-paced, electronic professional development may be more suited to the time restraints of principals. Therefore, districts, IUs, and universities should embrace these opportunities to attract more participants.

The key for districts and IUs is to find the balance between programs that offer collaboration with colleagues in order to share ideas as well as programs that can be flexibly delivered to meet the time constraints that principals experience.

Recommendations for Further Research

Based on the responses to this survey, future researchers may want to specifically target areas of low socio-economic status for a replication of this survey. Special interest would be in the Pittsburgh School District where principals reported in 2011 a lack of confidence in implementing the Framework (Lane & Horner, 2011). Allentown School District, which has gone through extensive professional development training regarding the implementation of the Framework (White, personal correspondence, August 2, 2014), would be another area of low socio-economic status which may want to measure the success of their professional training.

In light of the limitation previously discussed through self-report data, future research may try to determine the depth of principals' knowledge and competence in identifying specific evidence in evaluating the elements of the Framework. Although Teachscape provides a 20 – 25 hour on-line assessment of principals' ability to identify

evidence of the elements in a lesson, I was unable to ascertain the percentage of principals who have successfully completed the assessment. Mandatory recertification of principals requiring successful completion of the Teachscape assessment would be one way to determine the proficiency of principals' evaluations.

This research examined a principal's number of years of experience as a variable in relation to a principal's confidence in using the Framework. Further research may want to examine the relationship between the number of years a principal has served as a teacher prior to becoming a principal and the principal's efficacy in using the Framework. This may be particularly of interest as teachers begin to reflect on their teaching using the Framework as their guide for best instructional practices.

The logical next step following principals' ability to identify evidence or lack of evidence of the elements in a lesson would be for principals to be able to conduct meaningful conversations with teachers and offer constructive feedback. Because the topic of critical conversations is so important, I had planned to include a section regarding principals' confidence in conducting effective conferences in this survey. Unfortunately, the addition of this area of concern made the survey document too cumbersome. Future research should be done to determine the professional development needs of principals in the area of conducting critical conversations with teachers.

Finally, regarding professional development formats, future research may want to tease out and clarify the definitions of the professional development formats that principals prefer. For example, in this research, it is not clear whether it is Intermediate Units (Boards of Cooperative Educational Services, similar organizations outside the state of Pennsylvania) or district workshops that principals prefer. Also, there should be

more clarity between university course work and university sponsored professional development workshops.

Researchers may also want to investigate the electronic professional development opportunities that are now becoming popular. Days of traditional methods of professional development with instructors teaching and students sitting in a classroom are a thing of the past. Formats that weren't even considered for this research are being used to share ideas and strategies. Professional development of the future appears to be instantaneous through electronic devices, inexpensive, and specific to a principal's needs.

Conclusions

The purpose of this study was to determine the confidence levels of principals in implementing Danielson's Framework as a means to evaluate teachers. In addition, the study also identified principals' preferences for professional development formats. Findings from the study indicated that principals are relatively confident in evaluating teachers using the framework. It appeared that professional development opportunities would be beneficial in the area of identifying evidence where students take ownership of their learning and the learning environment. Using assessments in instruction would be another topic where principals could hone their skills.

Furthermore, school district and intermediate unit professional development opportunities were the preference for principals. This research challenges school districts and intermediate units to be on the cutting edge of educational theory in relation to teaching and learning in order to give principals the training needed to be 21st century leaders.

References

Anagnostopoulos, D., Rutledge, S. (2007). Making sense of school sanctioning policies in urban high schools. *Teachers College Record* , 109 (5), 1261–1302.

Anderson, J. B. (2008). Principals' role and public primary schools' effectiveness in four Latin American cities. *The Elementary School Journal,* 109 (1), 36-60.

Andrews, R. L., & Soder, R. (1987). Principal leadership and student achievement. *Educational Leadership*, 44(6), 9-11.

Barth, R. S. (1986). Principal centered professional development. *Theory into Practice*, 25(3), 156-160.

Bennett, A. (2002). Critical Issue: Guiding principals – Addressing accountability challenges. North Central Regional Education Laboratory. Retrieved April 2014 from http://www.ncrel.org/sdrs/areas/issues/educatrs/leadrship/le600.htm

Bill and Melinda Gates Foundation (2009). Retrieved June 8, 2014 from http://www.gatesfoundation.org/Media-Center/Press-Releases/2009/11/Foundation-Commits-$335-Million-to-Promote-Effective-Teaching-and-Raise-Student-Achievement

Bill and Melinda Gates Foundation (2010). Working with teachers to develop fair and reliable measures of effective teaching. Measures of Effective Teaching Project. Retrieved April 17, 2013 from http://www.metproject.org

Bill and Melinda Gates Foundation (2012). Gathering feedback for teaching. Measures of Effective Teaching Project. Retrieved April 17, 2013 from http://www.metproject.org

Bill and Melinda Gates Foundation (2013). Ensuring fair and reliable measures of

effective teaching. Measures of Effective Teaching Project. Retrieved April 17, 2013 from http://www.metproject.org

Blase, J. & Blase, J. (1999). Effective instructional leadership. *Journal of Educational Administration,* 38(2), 130-141.

Booher-Jennings, J. (2006). Rationing education in an era of accountability. *Phi Delta Kappan.* 85(10), 756-761.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268.

Branch, G. F., Hanushek, E. A., Rivkin, S. G., (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals*. NBER Working Paper No. 17803.

Brandt, R. (1987). On teacher evaluation: A conversation with Tom McGreal. *Educational Leadership*, 44(7), 20-24.

Bryant, K. A. (2011). Principal training, efficacy, and behavior in using student achievement data in the teacher evaluation process. *Proquest Information and Learning Company*. (UMI No. 3510040).

Camburn, E. M., Spillane, J. P., Sebastian, J. (2010). Assessing the utility of a daily log for measuring principal leadership practice. *Educational Administration Quarterly,* 46, 707-735.

Clifford, M, Behrstock-Sherratt, E., & Fetters, J. (2012). The ripple effect: A synthesis of research on principal influence to inform performance evaluation design. American Institutes for Research. www.air.org

Colby, S. A., Bradshaw, L. K., Joyner, R. L. (2002). Teacher evaluation: A review of

the literature. Paper presented at the Annual Meeting of the American Educational Research Association: New Orleans.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: Office of Education, U. S. Department of Health, Education and Welfare.

Corcoran, T. B. (1995). Helping teachers teach well: Transforming professional development. Consortium for Policy Research in Education. Policy Briefs.

Cotton, K. (2003). *Principals and student achievement: What the research says*. Alexandria, VA: Association for Supervision and Curriculum Development.

Council of Chief State School Officers. (2011). InTASC model core teaching standards: A resource for state dialogue. Retrieved June 20, 2013 from http://www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Stan2011.pdf

Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy, and practice at the district level. *Harvard Educational Review*, 54(2), 129-151.

Danielson, C. (2007). Enhancing professional practice: A framework for teaching. Association for Supervision and Curriculum Development, Alexandria, VA.

Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.

Danielson, C. (2011a). *Framework for teaching evaluation system*. Princeton, NJ: The Danielson Group.

Danielson, C. (2012). It's your evaluation – collaborating to improve teacher practice. *Education Digest*, 77(8), 22-27.

Danielson, C. and McGreal, T. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2012) The right start: Creating a strong foundation for the teaching career. *Phi Delta Kappan*, 94(3), 8-13.

Davis, S., Darling-Hammond, L., LaPointe, M., & Myerson, D. (2005). School leadership study: Developing successful principals. Stanford Educational Leadership Institute. www.wallacefoundation.org

Day, C., Sammons, P., Hopkins, D., Leithwood, K., & Kington, A. (2008). Research into the impact of school leadership on pupil outcomes: Policy and research contexts. *School Leadership and Management*, 28(1), 5-25.

DeBevoise, W. (1982). The principal's role. How do we reconcile expectations with reality? Washington, DC: National Institute of Education.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., Birman, B. F. (2002). Effects of professional development on teacher's instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81-112.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record.* 106(6), 1145–1176.

Donaldson, M. L. (2009). So long, lake Wobegon? Using teacher evaluation to raise teacher quality. Center for American Progress. Washington, DC.

Donaldson, M. L., Donaldson, G. A. (2012). Strengthening teacher evaluation what district leaders can do. *Educational Leadership*, 69(8), 78-82.

Duke, D. L. (1987). *School leadership and instructional improvement*. New York:

Random House.

Edmonds, R. (1982). *Programs of school improvement: An overview*. Paper presented at the National Invitational Conference. Warrenton, VA.

Ellett, C. D. and Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education* 17(1), 101-128.

Elmore, R. (2000). Building a new structure for leadership. Albert Shanker Institute. Washington, D.C.

Fink, E. & Resnick, L. B. (2001). Developing principals as instructional leaders. *The Phi Delta Kappan,* 82(8), 598-606.

Firestone, W. A. (2009). Accountability nudges districts into changes in culture. *The Phi Delta Kappan,* 90(9), 670-676.

Frase, L. E., (1992). Constructive feedback on teaching is missing. *Education*, 113(2), 176-181.

Gall, M. D., Gall, J. P., Borg, W. R. (2007). *Educational research: An introduction.* New York: Pearson Education, Inc.

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79-107.

Gentilucci, J., L. & Muto, C. C. (2007). Principals' influence on academic achievement: The student perspective. NASSP Bulletin, 91 (3), 219.

Glickman, C. D., Gordon, S. P., Ross-Gordon, J. M. (2001). *SuperVision and instructional leadership.* Needham Heights, MA: Allyn & Bacon.

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. National Comprehensive Center for Teacher Quality. Washington, D.C.

Goe, L. (2009). Measuring Teacher Effectiveness. National Comprehensive Center for Teacher Quality. Retrieved June 20, 2013 from http://www.gtlcenter.org/sites/default/files/docs/evaluateEffectiveness/Goe_TeacherEffectiveness.pdf

Goldrick, L. (2002). Improving teacher evaluation to improve teaching quality. Education Policy Studies Division, National Governors Association Center for Best Practices.

Grissom, J. A. & Harrington, J. R. (2010). Investing in administrator efficacy: An examination of professional development as a tool for enhancing principal effectiveness. *American Journal of Education*, 116(4), 583-612.

Haefele, D. L. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation In Education, 7*(4), 21-31.

Hallinger, P. (2003) Leading educational change: reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, 33(3), 329 — 352.

Hallinger, P. (2008). *Methodologies for studying school leadership: A review of 25 years of research using the Principal Instructional Management Rating Scale.* Paper prepared for presentation at the annual meeting of the American Educational Research Association. New York.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school

effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2), 157-191.

Hallinger, P., & Murphy, J. (1985). Assessing the instructional leadership behavior of principals. *Elementary School Journal, 86* (2), 217-248.

Hamilton, L. S., Engberg, J., Steiner, E. D., Nelson, C. A. & Yuan, K. (2010). Improving school leadership through support, evaluation, and incentives. The Pittsburgh Incentive Project. Retrieved June 8, 2014 from http://www.rand.org/pubs/monographs/MG1223.html

Hazi, H. M., Rucinski, D. A. (2009). Teacher evaluation as a policy target for improved student learning: A fifty-state review of statute and regulatory action since NCLB. *Education Policy Analysis Archives*, 17(5) 1-22.

Heck, R. H., & Hallinger, P. (2010). Collaborative Leadership Effects on School Improvement: Integrating Unidirectional- and Reciprocal-Effects Models. *The Elementary School Journal*, 111(2), 226-252.

Ho, A. D., Kane, T. J. (2013). The reliability of classroom observations by school personnel. Bill and Melinda Gates Foundation.

Holifield, M., Cline, D. (1997). Clinical Supervision and its outcomes: Teachers and principals report. NASSP Bulletin, 81(590), 109-113.

Holland, P. (2005). The case for expanding standards for teacher evaluation to include an instructional supervision perspective. *Journal of Personnel Evaluation in Education*, 18, 67-77.

Howard, B. B. and McCloskey, W. H. (2001). Evaluating experienced teachers.

Educational Leadership, 58(5), 48 – 51.

Jacob, B. A. & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics,* 26, 101-136.

Jaquith, A., Mindich, D., Wei, R. C., Darling-Hammond, L. (2011). Teacher professional learning in the U.S. *Education Digest*, 77(2), 33-40.

Jensen, R. & Moller, J. (2013). School data as mediators of professional development. *Journal of Educational Change*, 14, 95-112.

Joint Committee on Standards for Educational Evaluation. (1988). Retrieved June 8, 2014 from http://www.jcsee.org/program-evaluation-standards-statements

Kane, T. J., McCaffrey, D. F., Miller, T., Staiger, D. (2013). Have we identified effective teachers? Bill and Melinda Gates Foundation.

Kerrins, J. A. & Cushing, K. S. (2000). Taking a second look: Expert and novice differences when observing the same classroom teaching segment a second time. *Journal of Personnel Evaluation in Education*, 14(1), 5-24.

Kersten, T, & Israel, M. (2005). Teacher evaluation: Principals' insights and suggestions for improvement. *Planning and Changing*. 26(1&2), 47-67.

Kessinger, T. A. (2011). Efforts toward educational reform in the United States since 1958. A review of seven major initiatives. *American Educational History Journal*, 38(2), 263 – 276.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78.

Kraemer, H. C. and Thiemann, S. (1987). *How many subjects? Statistical power analysis in research.* Newbury, CA: Sage Publications, Inc.

Ladd, H. F. & Zelli, A. (2002). School-based accountability in North Carolina: The responses of school principals. *Educational Administration Quarterly,* 38(4), 494-529.

Lane, S. & Horner, C. (2011). Pennsylvania teacher and principal evaluation pilot final report. Team PA and PA Department of Education. Pittsburgh, PA.

Lashway, L. (2003). Transforming principal preparation. ERIC Clearinghouse on Educational Management. ERIC Digest (ED473360), 1-8.

Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly*, 30, 498-518.

Leithwood, K, (2005). Educational leadership. The Laboratory for Student Success at Temple University Center for Research in Human Development and Education. http://www.temple.edu/lss

Leithwood, K., & Duke, D. L. (1998). Mapping the conceptual terrain of leadership: A critical point of departure for cross-cultural. *Peabody Journal of Education*, 73(2), 31.

Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership and Management*, 28 (1), 27-42.

Leithwood, K., & Jantzi, D. (1998). *Distributed leadership and student engagement in school*. Paper presented at the Annual Meeting of the American Educational Research Association. San Diego, CA.

Leithwood, K., & Jantzi, D. (1999). *The effects of transformational leadership on*

organizational conditions and student engagement with school. Paper presented at the Annual Meeting of the American Educational Research Association. Montreal, Quebec, Canada.

Leithwood, K. & Jantzi, D. (2008**).** Linking leadership to student learning: The contributions of leader efficacy. *Educational Administration Quarterly.* 44 (4), 496-598.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004) *Review of research: how leadership influences student learning.* New York: The Wallace Foundation.

Levine, A. (2005). Educating school leaders. The Education Schools Project. Washington, DC.

Lipscomb, S., Chiang, H., Gill, B. (2012). Value-added estimates for phase 1 of the Pennsylvania teacher and principal evaluation pilot. Mathematica Policy Research. Cambridge MA.

Looney, J. (2011). Developing high-quality teachers: teacher evaluation for improvement. *European Journal of Education,* 46(4), 440-455.

Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39(3), 370-397.

Marshall, K. (2005). It's time to rethink teacher supervision and evaluation. *The Phi Delta Kappan*, 86(10), 727-735.

Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, 70(3), 14-19.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works*. Alexandria, VA: Association for Supervision and Curriculum Development.

McGhee, M. W., & Nelson, S. W. (2005). Sacrificing leaders, villainizing leadership: How educational accountability policies impair school leadership. *The Phi Delta Kappan*, 86(5), 367-372.

McGrath, M. J. (2000). The human dynamics of personnel evaluation. *The School Administrator.* October, 1-6.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. Peabody Journal of Education, 79(4), 33-53.

Milhaly, K., McCaffrey, D. F., Staiger, D. O., Lockwood, J. R. (2013). A composite estimator of effective teaching. MET Project. Bill and Melinda Gates Foundation. Retrieved April 17, 2013 from http://www.metproject.org

Mintrop, H. and Sunderman, G. L. (2009). Why high stakes accountability sounds good but doesn't work – and why we keep on doing it anyway. The Civil Rights Project/Proyecto Derechos Civiles at UCLA, Los Angeles, CA.

Mortimore, P. (1993). *School effectiveness and the management of effective learning and teaching.* Paper presented at The International Congress for School Effectiveness and Improvement. Norrkoping, Sweden.

Murphy, J. (1992). *School effectiveness and school restructuring: Contributions to educational improvement.* Paper presented at the Annual Meeting of the International Congress for School Effectiveness and Improvement. Victoria, British Columbia, Canada.

Nancarrow, C. and Brace, I. (2000). Saying the "right thing": Coping with social desirability bias in marketing research, *Bristol Business School Teaching and Research Review,* 3.

Nolan, J. F. and Hoover, L. A. (2004). *Teacher supervision and evaluation: Theory into practice.* Hoboken, NJ: John Wiley & Sons, Inc.

Norusis, M. J. (1994). *SPSS Professional Statistics 6.1*. Chicago, Ill: SPSS, Inc.

O'Brien, T. V. & Roberson, T. J. (2012). Not quite "deja vu all over again": No child left behind meets effective schools research, *The Educational Forum*, 76(3), 356-371.

Odden, A. (2004). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education*, 79(4), 126-137.

Ovando, M. N. (1994). Constructive feedback: A key to successful teaching and learning. *International Journal of Educational Management*, 8(6), 19–22.

Ovando, M. N. (2006). Building instructional leaders' capacity to deliver constructive feedback to teachers. *Journal of Personnel Evaluation in Education,* 18, 171-183.

Pajak, E. (1990). Dimensions of supervision. *Educational Leadership,* 48(1), 78-81.

Painter, S. R. (2000). Principals' perceptions of barriers to teacher dismissal. *Journal of Personnel Evaluation in Education*, 14(3), 253-264.

Painter, S. R. (2000a). Principals' efficacy beliefs about teacher evaluation. Journal of Educational Administration. 38(4), 368-378.

Papay, J. P. (2012). Refocusing the Debate: Assessing the Purposes and Tools of Teacher Evaluation. *Harvard Educational Review,* 82(1), 123 – 167.

Pennsylvania Department of Education (2013a). Educator effectiveness administrative manual September 2013 – Part I. Retrieved October 3, 2013 from

http://www.portal.state.pa.us/portal/server.pt/community/educator_effectiveness_project/20903

Pennsylvania Department of Education (2013b). Educator effectiveness program. Retrieved June 20, 2013 from http://www.portal.state.pa.us/portal/server.pt/community/educator_effectiveness_project/20903

Pennsylvania Department of Education (2013c). Data and Statistics. Retrieved April 18, 2015 from http://www.portal.state.pa.us/portal/server.pt/community/professional_and_support_personnel/7429

Pennsylvania Department of Education (2013d). Data and Statistics. Retrieved April 18, 2015 from http://www.portal.state.pa.us/portal/server.pt/community/professional_and_support_personnel

Pennsylvania Department of Education (2013e). Data and Statistics. Retrieved April 22, 2015 from http://paschoolperformance.org/SelectCounty

Pennsylvania General Assembly. Public School Code of 1949 - Omnibus Amendments. Act of Jun. 30, 2012, P.L. 684, No. 82 retrieved February 27, 2013 from http://www.legis.state.pa.us/WU01/LI/LI/US/HTM/2012/0/0082..HTM

Pennsylvania State Education Association (2012). Teacher evaluation and race to the top in Pennsylvania. Retrieved June 20, 2013 from http://www.psea.org/uploadedFiles/Publications/Professional_Publications/Advisories/Advisory-TeacherEvaluation-RTTT-April2012.pdf

Persell, C., Cookson, P., & Lyons, H. (1982). Effective principals: What do we know

from various educational literatures? National Institute of Education. Washington, D.C.

Peterson, K. (2002). The professional development of principals: Innovations and opportunities. *Educational Administration Quarterly*, 38(2), 213-232.

Pitner, N. J. (1988). The study of administrator effects and effectiveness. In N. J. Boyan (Ed.), *Handbook of research on educational administration* (99 – 122). New York: Longman.

Porter, A. (2000). Doing high-stakes assessment right. *School Administrator*, 57(11), 28-31.

Portin, B., Schneider, P., DeArmond, M., & Gundlach, L. (2003). Making sense of leading schools: A study of the school principalship. Center on Reinventing Public Education. www.wallacefoundation.org

Principal Instructional Management Rating Scale. Retrieved May 19, 2014 from http://alex.state.al.us/leadership/Principals%20%20Files/I-12,%20Hallinger,%20PMRS%20Teacher%20Form11_13_07.pdf

Ravitch, D. (2011). Dictating to the schools. *Education Digest*, 76(8), 4-9.

Rice, J. K. & Malen, B. (2003). The human costs of education reform: The case of school reconstitution. *Educational Administration Quarterly*, 39(5), 635-666.

Roberson, T. S. (2010). *The relationship between principal personality type and elementary school student achievement.* Unpublished doctoral dissertation, Walden University, Minneapolis, MN.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types.

Educational Administration Quarterly, 44(5), 635-674.

Rudner, L. M., Glass, G. V., Evartt, D. L., & Emery, P. J. (2002). A user's guide to the meta-analysis of research studies. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Rutledge, S., Harris, D., & Ingle, W. (2010). How principals "bridge and buffer" the new demands of teacher quality and accountability: A mixed-methods analysis of teacher hiring. *American Journal of Education,* 116, 211-242.

Sanders, W. L. & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. University of Tennessee Value-Added Research and Assessment Center. Knoxville, TN.

Saphier, J. (2011). Outcomes. *Journal of Staff Development*, 33(4), 58-62.

Sartain, L., Stoelinga, S. R., Brown, E. (2011). Rethinking teacher evaluation in Chicago. Research Report. Consortium on Chicago School Research at the University of Chicago Urban Education Institute. Chicago, Illinois.

Sawyer, L. (2001). Revamping a teacher evaluation system. *Educational Leadership*, 58(5), 44-47.

School District of Philadelphia (2015). About Us. School Meals Served Daily. Retrieved July 21, 2014 from http://www.phila.k12.pa.us/about/#school-meals

Seashore-Louis, K., Leithwood, K., Wahlstrom, K. L., Anderson, S. E. (2010). Investigating the links to improved student learning. Center for Applied Research and Educational Improvement. www.wallacefoundation.org

Sergiovanni, T. J. (2006). *The principalship: A reflective practice perspective*. Boston,

MA: Pearson.

Sergiovanni, T. J. & Starratt, R. J. (1998). *Supervision: A redefinition*. Boston, MA: McGraw Hill.

Shakman, K., Breslow, N., Kochanek, J., Riordan, J. & Haferd, T. (2012). Changing cultures and building capacity:An exploration of sidtrict strategies for implementation of teacher evaluation systems. Retrieved June 8, 2014 from https://secure.edc.org/publications/prodview.asp?2124

Sheppard, B., Hurley, N., & Dibbon, D. (2010). *Distributed leadership, teacher morale, and teacher enthusiasm: Unraveling the leadership pathways to school success.* Paper presented at American Educational Research Association, Denver, CO.

Shoemaker, J., & Fraser, H. W. (1981). What principals can do: Some implications from studies of effective schooling. *The Phi Delta Kappan*, 63(3), 178-182.

Silva, J. (2010). The direct effects of principal-student discussions upon eighth grade students' gains in reading achievement: An experimental study. *Proquest Information and Learning Company*. (UMI No. 3404114).

Smith, W. F., & Andrews, R. L. (1989). *Instructional leadership: How principals make a difference.* Alexandria, VA: Association for Supervision and Curriculum Development.

Spillane, J. P., Kenney, A. W., (2012). School administration in a changing education sector: the US experience. *Journal of Educational Administration,* 50 (5), 541-561.

Stecher, B., Garet, M., Holtzman, D., & Hamilton, L. (2012). Implementing measures of teacher effectiveness. *Phi Delta Kappan*, 94(3), 39-43.

Stedman, Lawrence C. (1987). It's Time We Changed the Effective Schools Formula. *Phi Delta Kappan, 69*(3), 215.

Strizek, G., Pittsonberger, J., Riordan, K., Lyter, D., and Orlofsky, G. (2006). Characteristics of schools, districts, teachers, principals, and school libraries in the United States: 2003-04 schools and staffing survey (NCES 2006-313 Revised). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Sweeney, J. (1982). Research on effective school leadership. *Educational Leadership*, 39(5), 346-352.

Taylor, E. S. & Tyler, J. H. (2011). The Effect of evaluation on performance: Evidence from longitudinal student achievement data and mid-career teachers. Working paper from the National Bureau of Economic Research, Cambridge: MA.

Teachscape (2013) FOCUS for Observers. Retrieved February 20, 2014 at http://www.teachscape.com/products/focus/how-it-works/for-observers

Team Pennsylvania Foundation (2011a). Teacher evaluations highlighted in governor's education agenda. Published online October 11, 2011. Harrisburg, PA Retrieve from http://teampa.com/2011/10/teacher-evaluations-highlighted-in-governors- education-agenda/

Team Pennsylvania Foundation (2011b). Team PA, groups agree: PA needs a statewide teacher evaluation system. Published online June 8, 2011. Harrisburg, PA Retrieved from http://teampa.com/2011/06/its-unanimous-pa-needs-a-statewide-teacher-evaluation- system/

The National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Retrieved June 22, 2013 at

http://www2.ed.gov/pubs/NatAtRisk/index.html

Toch, T. & Rothman, R. (2008). Rush to judgment: Teacher evaluation in public education. Education Sector. Washington, D.C.

U.S. Congress, Senate Committee on Equal Educational Opportunity. (1970). Toward equal educational opportunity. Washington, DC: Government Printing Office.

U.S. Department of Education. (2002). *No Child Left Behind* (Public Law 107–110). Washington, DC: Author. Retrieved June 22, 2013 from http://www2.ed.gov/policy/elsec/leg/esea02/beginning.html

U.S. Department of Education (2009). Race to the Top Program. Executive Summary. Retrieved August 28, 2013 from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2012). Findings from the 2011-12 survey on the use of funds Under Title II, Part A. Retrieved September 20, 2013 from www2.ed.gov/programs/teacherqual/2012findings.doc

Vasquez-Heilig, J., Young, M., Williams, A. (2011). At-risk student averse: risk management and accountability. *Journal of Educational Administration*, 50(5), 562-585

Volkman, D. (2013) June 20, 2013 - Teachscape *Focus* Licenses for Evaluators. From RTT3 Communications Archive. Retrieved February 20, 2014 from https://search.yahoo.com/search;_ylt=AhW1iZ1XczrqrLTos0ormhWbvZx4?p=RTT3+Communications+Archive&toggle=1&cop=mss&ei=UTF-8&fr=yfp-t-901

Wahlstrom, K. L., Louis, K. S., Leithwood, K., Anderson, S. E. (2010). Investigating the links to improved student learning. The Wallace Foundation.

www.wallacefoundation.org

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. The New Teacher Project. www.tntp.org

Wiggins, G. (2012). 7 keys to effective feedback. *Educational Leadership*, 70(1), 10-16.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., and H. T., Bernstein. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal*, 86(1), 60-121.

Witziers, B., Bosker, R. J., & Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 389-425.

Appendix A

PDE 428

Commonwealth of Pennsylvania	DEPARTMENT OF EDUCATION	333 Market St., Harrisburg, PA 17126-0333

EMPLOYEE EVALUATION FORM FOR INSTRUCTIONAL II TEACHERS

Employee's Last Name | First | Middle | Positions(s) of Employee

District/IU | School | Evaluator | Interview/Conference Date

School Year: | Evaluation: (Check 1) ☐ One ☐ Two

This form is to serve as a permanent record of an administrator's evaluation of a teacher's performance during a specific time period based on specific criteria.

PERFORMANCE EVALUATION

Directions: Examine all sources of evidence provided by the teacher and bear in mind the aspects of teaching for each of the four categories used in this form. Refer to the rubric language, checking the appropriate aspects of teaching, and indicating the sources of evidence used to determine the evaluation of the results in each category. Last, assign an overall evaluation of performance, sign the form and gain the signature of the employee.

Category I: Planning and Preparation -- **Through their knowledge of content and pedagogy skills in planning and preparation, teachers make plans and set goals based on the content to be learned, their knowledge of students and their instructional context. Category I reviews: Knowledge of Content and Pedagogy, Knowledge of Students, Selecting Instructional Goals, Designing Coherent Instruction, Assessing Student Learning, Knowledge of Resources, Materials and Technology.**

☐ SATISFACTORY	☐ UNSATISFACTORY
Performance Demonstrates: o In-depth and thorough knowledge of content, pedagogy and Pa. Academic Standards o Thorough knowledge of students and how to use this knowledge to direct and guide instruction o Clear and appropriate instructional goals that reflect Pa. standards and high expectations for students o In-depth and thorough awareness of resources, materials, or technology available through the school or district or professional organizations o Appropriate instructional design in which plans for various elements are aligned with the instructional goals and have a recognizable sequence and required adaptations for individual student needs o Appropriate assessments of student learning completely aligned to the instructional goals and adapted as needed for student needs.	Performance Demonstrates: o Limited or partial knowledge of content, pedagogy and Pa. Academic Standards o Irrelevant or partial knowledge of students and how to use this information to direct and guide instruction o Unclear or trivial instructional goals and low expectations for students. o Little or no awareness of resources, materials, and technology available through the school or district or professional organizations o Inappropriate or incoherent instructional design in which plans for elements are not aligned with the instructional goals, and have few or inappropriate adaptations for individual student needs o Inappropriate assessments of student learning not aligned to the instructional goals nor adapted as needed for student needs.

Sources of Evidence (Check all that apply and include dates, types/titles and number)

☐ Lesson/Unit Plans	See Attachment 428 A	☐ Teacher Conferences/Interviews	See Attachment 428 A
☐ Resources/Materials/Technology	See Attachment 428 A	☐ Classroom Observations	See Attachment 428 A
☐ Assessment Materials	See Attachment 428 A	☐ Teacher Resource Documents	See Attachment 428 A
☐ Information About Students	See Attachment 428 A	☐ Other	See Attachment 428 A

Justification for Evaluation

Category II: Classroom Environment -- **Teachers establish and maintain a purposeful and equitable environment for learning, in which students feel safe, valued, and respected by instituting routines and by setting clear expectations for student behavior. Category II reviews: Teacher Interaction with Students, Establishment of an Environment for Learning, Student Interaction.**

☐ SATISFACTORY	☐ UNSATISFACTORY
Performance demonstrates: o High and clear expectations for student achievement with value placed on the quality of student work o Significant attention to equitable learning opportunities for students o Appropriate and highly respectful interactions between teacher and students and among students o Highly effective classroom routines and procedures resulting in little or no loss of instructional time o Clear standards of conduct and effective management of student behavior o Safe and skillful organization of physical space, to the extent it is under the control of the teacher, that provides accessibility to learning and to the use of resources.	Performance Demonstrates: o Low or unclear expectations for student achievement with little or no value placed on the quality of student work. o Little or no attention to equitable learning opportunities for students o Inappropriate or disrespectful interactions between teacher and students and among students o Inefficient classroom routines and procedures resulting in loss of instructional time o Absent or unclear standards of conduct, or ineffective management of student behavior o Unsafe or inadequate organization of physical space, to the extent it is under the control of the teacher, to provide accessibility to learning and to the use of resources, materials, and technology.

Sources of Evidence (Check all that apply and include dates, types/titles, and number)

☐	Classroom Observations	See Attachment 428 A	☐	Visual Technology	See Attachment 428 A
☐	Informal Observation/Visits	See Attachment 428 A	☐	Resources/Materials/Technology/Space	See Attachment 428 A
☐	Teacher Conferences/Interviews	See Attachment 428 A	☐	Other	See Attachment 428 A

Justification for Evaluation

Category III: Instructional Delivery -- **Through their knowledge of content and their pedagogy and skill in delivering instruction, teachers engage students in learning by using a variety of instructional strategies. Category III addresses: Communications, Questioning and Discussion Techniques, Engaging Students in Learning, Providing Feedback, Demonstrating Flexibility and Responsiveness**

☐ SATISFACTORY	☐ UNSATISFACTORY
Performance Demonstrates: o Clear and appropriate communication of procedures and high quality explanations of content o Highly effective use of questioning and discussion strategies that encourage many students to participate o High-level engagement of students in learning and adequate pacing of instruction o Equitable, accurate and constructive feedback to students on their learning o Informed and appropriate use of informal and formal assessments to meet learning goals and to monitor student learning o High degree of flexibility and responsiveness in meeting the learning needs of students.	Performance Demonstrates: o Unclear or inappropriate communication of procedures and poor explanations of content o Ineffective use of questioning and discussion strategies and little student participation o Little or no engagement of students in learning and poor pacing of instruction o Inaccurate or inappropriate feedback to students on their learning o Little or inappropriate use of formal and informal assessments to meet learning goals and to monitor student learning o Inflexibility in meeting the learning needs of students.

Sources of Evidence (Check all that apply and include dates, types/titles, or number)

☐	Classroom Observations	See Attachment 428 A	☐	Student Assignment Sheets	See Attachment 428 A
☐	Informal Observations/Visits	See Attachment 428 A	☐	Student Work	See Attachment 428 A
☐	Assessments Materials	See Attachment 428 A	☐	Instructional Resources/Materials/Technology	See Attachment 428 A
☐	Teacher Conferences/Interviews	See Attachment 428 A	☐	Other	See Attachment 428 A

Justification for Evaluation

Category IV: Professionalism – **Professionalism refers to those aspects of teaching that occur in and beyond the classroom/building. Category IV addresses: Adherence to School and District Procedures, Maintaining Accurate Records, Commitment to Professional Standards, Communicating with Families, Demonstrating Professionalism.**

☐ SATISFACTORY	☐ UNSATISFACTORY
Performance Demonstrates: o Full adherence to school and district procedures and regulations related to attendance, punctuality, and the like. o Full knowledge of Professional Code of Conduct and full commitment to professional standards o Full and active compliance with school and district requirements for maintaining accurate and complete records o Full and active compliance with district requirements for communicating with families regarding student needs/improvement o Full and frequent participation in professional development events/opportunities, consistent application of new learning in the classroom, and sharing of learning with colleagues	Performance Demonstrates: o Little and/or irregular compliance to school and district procedures and regulations related to attendance, punctuality, and the like o Little knowledge of Professional Code of Conduct and little commitment to professional standards o Inefficient or ineffective system for maintaining accurate records that is not in compliance with school or district guidelines o Infrequent or inappropriate communication with families to understand student needs and development o Little or infrequent participation in professional development opportunities, little application of new learning in the classroom and little sharing of learning with colleagues

Sources of Evidence (Check all that apply and include dates, types/titles, or number)

☐	Teacher Conferences/Interviews	See Attachment 428 A	☐	Progress Reports/Report Cards	See Attachment 428 A
☐	Observations/Visual Technology	See Attachment 428 A	☐	Parent/School/Community Feedback	See Attachment 428 A
☐	Artifacts/Interaction with Family	See Attachment 428A	☐	Artifacts: Professional Development/Act 48 Documentation	See Attachment 428 A
☐	Student Records/Grade Book	See Attachment 428 A	☐	Perceptive Use of Teaching/Learning Reflections	See Attachment 428 A
			☐	Other	See Attachment 428 A

Justification for Evaluation

Commonwealth of Pennsylvania	DEPARTMENT OF EDUCATION	333 Market St., Harrisburg, PA 17126-0333

Evaluation:
I certify that the before named employee for the period beginning (month/day/year) and ending (month/day/year) has been evaluated with a overall assessment that is: ☐Satisfactory ☐Unsatisfactory ☐Commendable

Signature of Principal/Assistant Principal (Evaluator) Date

Signature of Superintendent or I. U. Director Date

Overall Justification for Evaluation

Commendations (optional)

Professional Development Areas:

Name of Employee **Signature of Employee** **Date**

Appendix B

PA 82-1

Commonwealth of Pennsylvania — 333 Market St., Harrisburg, PA 17126-0333

pennsylvania DEPARTMENT OF EDUCATION

CLASSROOM TEACHER RATING TOOL FORM

Last Name	**First**	**Middle**
District/LEA	**School**	
Rating Date:	**Evaluation: (Check one)**	**Semi-annual** **Annual**

(A) Teacher Observation and Practice

Domain	Title	*Rating* (A)	Factor (B)	Earned Points (A x B)	Max Points
I.	Planning & Preparation		20%		0.60
II.	Classroom Environment		30%		0.90
III.	Instruction		30%		0.90
IV.	Professional Responsibilities		20%		0.60
(1)Teacher Observation & Practice Rating					3.00

Domain Rating Assignment 0 to 3 Point Scale (A)	
Rating	**Value**
Failing	0
Needs Improvement	1
Proficient	2
Distinguished	3

(B) Student Performance - Building Level Data, Teacher Specific Data, and Elective Data

Building Level Score (0 – 107)		**(3) Teacher Specific Rating**	
(2) Building Level Score Converted to 3 Point Rating		**(4) Elective Rating**	

(C) Final Teacher Effectiveness Rating – All Measures

Measure	Rating (C)	Factor (D)	Earned Points (C x D)	Max Points
(1) Teacher Observation & Practice Rating		50%		1.50
(2) Building Level Rating		15%		0.45
(3) Teacher Specific Rating		15%		0.45
(4) Elective Rating		20%		0.60
Total Earned Points				3.00

Conversion to Performance Rating	
Total Earned Points	**Rating**
0.00-0.49	Failing
0.50-1.49	Needs Improvement
1.50-2.49	Proficient
2.50-3.00	Distinguished
Performance Rating	

☐ **Rating: Professional Employee,** ***OR*** ☐ **Rating: Temporary Professional Employee**

I certify that the above-named employee for the period beginning _____ and ending _____ has received a performance rating of:
(month/day/year) (month/day/year)

☐ **DISTINGUISHED** ☐ **PROFICIENT** ☐ **NEEDS IMPROVEMENT** ☐ **FAILING**

resulting in a FINAL rating of:

☐ **SATISFACTORY** ☐ **UNSATISFACTORY**

A performance rating of Distinguished, Proficient or Needs Improvement shall be considered satisfactory, except that the second Needs Improvement rating issued by the same employer within 10 years of the first final rating of Needs Improvement where the employee is in the same certification shall be considered unsatisfactory. A rating of Failing shall be considered unsatisfactory.

_____ _____ _____ _____
Date Designated Rater / Position: Date Chief School Administrator

I acknowledge that I have read the report and that I have been given an opportunity to discuss it with the rater.
My signature does not necessarily mean that I agree with the performance evaluation.

Appendix C

Pilot Study Invitation

Dear Colleague,

I am the principal of Doyle Elementary School, Central Bucks School District as well as a doctoral student at Lehigh University under the supervision of Dr. George White. My dissertation study focuses on the readiness of principals as instructional leaders to implement the Danielson Framework as part of the Pennsylvania Educator Effectiveness System. It is my goal to determine the staff development needs of principals in order for principals to effectively implement the new evaluation system.

The purpose of this letter is to request your participation in a pilot study of my dissertation questionnaire. I am looking for your feedback regarding the design of the survey, clarity of the questions, readability, time it takes to complete the survey, and any other information you may offer to refine the questionnaire. All information from the pilot study will be confidential and will not be part of the study.

Below you will find the link to the questionnaire that contains all of the questions that will be part of the study as well as reflection questions for you answer to supply feedback to me on specific aspects of the questionnaire. When you click on the link, please time your response to the questionnaire, without including the feedback questions. Answer all questions. The information you share will be registered anonymously and will be returned electronically to me.

If you have any questions about my research, please feel free to contact me at Doyle Elementary School, (267)-893-4349, by cell (215) 778-3824, or email at ssalvesen@cbsd.org. You may also contact my dissertation advisor Dr. George White at Lehigh University, (610) 758-3262 or gpw1@lehigh.edu.

Click on the link:

https://www.surveymonkey.com/s/85J98QK

to access the questionnaire. Please complete the survey by no later than January 23.

Thank you in advance for your time and professional feedback.

Sincerely,

Susan L. Salvesen, Principal
Doyle Elementary School
Doctoral Student, Lehigh University

Appendix D

Principal Professional Development Pilot Scale

Part 1 Danielson's Framework

Please indicate whether you agree or disagree with each of the following statements pertaining to assessment of the elements of Components 2 and 3 of Danielson's Framework.

All responses are:

1. Strongly Agree 2. Agree 3. Disagree 4. Strongly Disagree

Each statement begins with:

I am able to confidently align the evidence to assess…

1. the lesson adjustment when needed.
2. student interactions with other students, including both words and actions.
3. assessment feedback to students.
4. the culture of high expectations for learning and achievement.
5. the assessment criteria.
6. the management of instructional groups.
7. the engagement of students using instructional materials and resources.
8. the communication of content explanations.
9. the engagement of students in learning through activities and assignments.
10. behavioral expectations.
11. teacher and student discussion techniques.
12. the response to student misbehavior.
13. the use of oral and written language with students.
14. the communication of directions and procedures.

15. the quality of questions and prompts.

16. the management of materials and supplies.

17. the monitoring of student behavior.

18. student participation.

19. the performance of non-instructional duties.

20. the grouping of students.

21. the management of transitions.

22. the structure and pacing of the lesson.

23. students' pride in their work.

24. monitoring of student learning through assessment.

25. the teacher's ability to convey the importance of the content and of learning.

26. the students' self-assessment and monitoring of progress.

27. teacher interactions with students, including both words and actions.

28. the communication of expectations for learning.

Part 2 Professional Development Formats

How likely or unlikely would you be to pursue professional development in these formats?

All responses for 29 - 37 are:

1. Very Likely 2. Likely 3. Unlikely 4. Very Unlikely

29. District or IU sponsored workshops

30. Independently sponsored self-paced online courses

31. Mentoring and/or coaching sessions

32. University course work on campus

33. University course work hybrid: on campus and partially online

34. University course online

35. Small study groups

36. State or National Conferences

37. Are there other professional development formats you would like to pursue?

_____ yes

_____ no

If yes, please describe them here _____

Part 3 Individual and Organizational Characteristics

38. What level is your school? Elementary, Middle, High

39. What percentage of free and reduced lunch students are in your school? _____

40. What was your school's most recent School Performance Profile percentage? _____

41. How many years of experience do you have as a principal? _____

42. Do you use the Danielson's Framework as a basis for your teacher evaluations?

Yes, No

If No, please identify your model _____

43. On average, how many teachers do you evaluated each year? _____

Time required to complete the survey _____ minutes

Are the directions for completing the survey clear? _____

If no, please offer suggestions for improvement (open-ended)

Are the questions in Part 1 Danielson's Framework clear? _____

If no, please identify the questions by number and then offer specific suggestions for improvement (open-ended)

Was there enough detail in the questions regarding the Framework to allow you to make a meaningful differentiation between agree and disagree? Yes/No

Are the questions in Part 2 Professional Development clear? _____

If no, please identify the questions by number and then offer specific suggestions for improvement (open-ended)

Are the questions in Part 3 Individual and Organizational Demographics clear?____

If no, please offer suggestions for question improvement (open-ended)

Other suggestions (open-ended)

Appendix E

Superintendent Support Letter

Dear Superintendent _____,

I am the principal of Doyle Elementary School, Central Bucks School District as well as a doctoral student at Lehigh University under the supervision of Dr. George White. The purpose of this letter is to request your support in having your principals participate in my dissertation research.

The focus of my dissertation is to examine the readiness of principals as instructional leaders to implement the Danielson Framework as part of the Pennsylvania Educator Effectiveness System. It is my goal to determine the staff development needs of principals in order for principals to effectively implement the new evaluation system. Principal participation is completely voluntary and anonymous. There are no foreseeable risks to participation in this study.

In order to have the necessary response rate to make the research meaningful, I am asking for your assistance. It would be helpful if you would acknowledge to your administrative team that this survey is being conducted and encourage your principals to complete the questionnaire in a timely manner. As a token of my appreciation for your principals' participation, I will provide you with a summary of the study's findings upon completion of the project.

Principals in your district will receive an electronic survey. The information gathered from the survey will provide me with data to determine the professional development needs of principals. Being respectful of their time, the entire questionnaire is less than 45 questions and results from a pilot study show that completion time averages less than 9 minutes.

If you have any questions about my research, please feel free to contact me at Doyle Elementary School, (267)-893-4349, by cell (215) 778-3824, or email at sls205@lehigh.edu. You may also contact my dissertation advisor Dr. George White at Lehigh University, (610) 758-3262 or gpw1@lehigh.edu.

Thank you for your time and support.

Sincerely,
Susan L. Salvesen, Principal
Doyle Elementary School
Doctoral Student, Lehigh University

Appendix F

Participant Invitation

Dear Colleague,

I am the principal of Doyle Elementary School in the Central Bucks School District. I am also a doctoral student in Educational Leadership at Lehigh University. The purpose of this letter is to request your participation in my dissertation research.

The focus of my dissertation is to examine the readiness of principals as instructional leaders to implement the Danielson Framework as part of the Pennsylvania Educator Effectiveness System. It is my goal to determine the staff development needs of principals in order for principals to effectively implement the new evaluation system.

Your assistance is needed in providing information through the completion of a brief survey. Participation in this survey is voluntary. In consideration of your busy schedule, results from a pilot study show that completion time averages less than 9 minutes.

All individual responses will remain anonymous, and only aggregate data will be presented. Completion and submission of this survey will constitute consent to participate. There are no known risks to participation. As an incentive to participate in this study, all participating principals may choose to enter into a drawing to win 15 prizes of $20 Amazon.com gift cards. Contact information for the drawing will be confidential.

Your help in completing this survey will provide professional development trainers in the field of Educational Leadership valuable information regarding the professional development needs of principals in relation to the Danielson Framework. First, specific components of the Danielson rubric will be identified as areas where principals perceive a need for professional growth. Second, instructors will receive insights into the formats principals prefer for the delivery of this professional learning.

Click on the link:

https://www.surveymonkey.com/s/RL2T8BC

to access the questionnaire. Please complete the questionnaire by February 27, 2015.

If you have any questions about my research, please feel free to contact me at Doyle Elementary School, (267)-893-4349, by cell (215) 778-3824, or email at ssalvese@cbsd.org. You may also contact my dissertation advisor Dr. George White at Lehigh University, (610) 758-3262 or gpw1@lehigh.edu. or Susan Disidore of the Office of Research Integrity at sus5@lehigh.edu.

Thank you for your time and support.

With sincere appreciation,
Susan L. Salvesen, Principal
Doyle Elementary School
Doctoral Student, Lehigh University

Appendix G

Principal Professional Development Scale

Part 1 Danielson's Framework

Please indicate whether you agree or disagree with each of the following statements pertaining to assessment of the elements of Components 2 and 3 of Danielson's Framework. All responses are:

1. Strongly Agree 2. Agree 3. Slightly Agree 4. Slightly Disagree 5. Disagree 6. Strongly Disagree

Note: these scores were reversed for purposes of analysis so that higher scores indicated greater agreement.

Each statement begins with:

In order to conduct **meaningful conversations** with teachers, I am able to **confidently** align the evidence to **assess**…

1. the lesson adjustment when needed.
2. student interactions with other students, including both words and actions.
3. assessment feedback to students.
4. the culture of high expectations for learning and achievement.
5. the assessment criteria.
6. the management of instructional groups.
7. the engagement of students using instructional materials and resources.
8. the communication of content explanations.
9. the engagement of students in learning through activities and assignments.
10. behavioral expectations.
11. teacher and student discussion techniques.
12. the response to student misbehavior.
13. the use of oral and written language with students.

14. the communication of directions and procedures.

15. the quality of questions and prompts.

16. the management of materials and supplies.

17. the monitoring of student behavior.

18. student participation.

19. the performance of non-instructional duties.

20. the grouping of students.

21. the management of transitions.

22. the structure and pacing of the lesson.

23. students' pride in their work.

24. monitoring of student learning through assessment.

25. the teacher's ability to convey the importance of the content and of learning.

26. the students' self-assessment and monitoring of progress.

27. teacher interactions with students, including both words and actions.

28. the communication of expectations for learning.

Part 2 Professional Development Formats

How likely or unlikely would you be to pursue professional development in these formats?

All responses for 29-37 are:

1. Very Likely 2. Likely 3. Unlikely 4. Very Unlikely
Note: these scores were reversed for purposes of analysis, so that higher scores indicated greater liklihood.

29. District or IU sponsored workshops

30. Independently sponsored self-paced online courses

31. Mentoring and/or coaching sessions

32. University course work on campus

33. University course work hybrid: on campus and partially online

34. University course online

35. Small study groups

36. State or National Conferences

37. Are there other professional development formats you would like to pursue?

_____ yes

_____ no

If yes, please describe them here _____

_____

Part 3 Individual and Organizational Characteristics

38. What level is your school? Elementary, Middle, High

39. What percentage of free and reduced lunch students are in your school? _____

40. What was your school's most recent School Performance Profile percentage? _____

41. How many years of experience do you have as a principal? _____

42. Do you use the Danielson's Framework as a basis for your teacher evaluations?

Yes, No

If No, please identify your model.

43. On average, how many teachers have you evaluated each year? _____

Appendix H

Participant Follow Up E-Mail

Dear Colleague,

Recently, I invited you to take part in an online questionnaire regarding principals' needs for professional development on the Danielson Framework. At this time, 188 principals have completed the survey. If you are one of these principals, thank you for your time and feedback. As a principal myself, I know how hectic your day can be. If you have not yet had time to fill in the survey, I invite you again to complete the questionnaire at:

https://www.surveymonkey.com/s/RL2T8BC

Your assistance is needed in providing information through the completion of a brief survey. Participation in this survey is voluntary. In consideration of your busy schedule, completion of this survey should take approximately 9 minutes. All individual responses will remain anonymous, and only aggregate data will be presented. Completion and submission of this survey will constitute consent to participate. There are no known risks to participation. As an incentive to participate in this study, all participating principals may choose to enter into a drawing to win 15 prizes of $20 Amazon.com gift cards. Contact information for the drawing will be confidential.

All information from participants will be kept confidential. If you have any questions about my research, please feel free to contact me at Doyle Elementary School, (267)-893-4349, by cell (215) 778-3824, or email at ssalvese@cbsd.org. You may also contact my dissertation advisor Dr. George White at Lehigh University, (610) 758-3262 or gpw1@lehigh.edu or Susan Disidore of the Office of Research Integrity at sus5@lehigh.edu.

Thank you for your time and support.

With sincere appreciation,

Susan L. Salvesen, Principal
Doyle Elementary School
Doctoral Student, Lehigh University

Susan L. Salvesen

PO Box 1148
Buckingham, PA 18912
215-794-7264
ssalvesen@cbsd.org

PROFESSIONAL PROFILE

A building level administrator with experience in strategic planning, curriculum development, teacher supervision, school improvement planning, professional community building, and community partnerships.

EDUCATION

Lehigh University
Doctor of Education, 2015
Concentration: Educational Leadership

Lehigh University
Superintendent Letter of Eligibility, 2012

Pennsylvania State University
Principal Certification, 1999

Wagner College
Master of Science, 1976
Concentration: Elementary Education

Wagner College
Bachelor of Science, 1972
Concentration: Elementary Education

PROFESSIONAL EXPERIENCE

2002 – Present Principal, Doyle Elementary School
Central Bucks School District

- Redesigned the Instructional Support Team
- Developed a new behavior plan
- Introduced backward planning through Understanding by Design
- Introduced the study of rigor, relevance, relationships, and reflection
- Established partnerships with the Doylestown Police Department, Doylestown Hospital, Delaware Valley University, community businesses, and cluster schools

2000 – 2002 Assistant Principal, Groveland Elementary School
Central Bucks School District

- Facilitated the opening of a new elementary school for 900 students, including: creating a mission statement, goals, and building philosophy, bringing families together from different schools, developing teams of teachers, and designing the interior décor of the building

1999 – 2000 Curriculum Coordinator for Social Studies and Science
Central Bucks School District

- Disseminated science materials for new science curriculum
- Ordered the curriculum materials for the opening of two new schools

1993 – 1999 6th Grade Teacher, Doyle Elementary School
Central Bucks School District

- Developed the Clean Stream curriculum in partnership with Peace Valley Nature Center, solicited funding from the School Board
- Presented Authentic Assessments to the School Board
- Promoted and solicited funds for the building of the Bike and Hike Path

1972 – 1979 7th Grade Science Teacher, Intermediate School 51
New York City School District

- Piloted a new science curriculum

AWARDS

2012 - 2013 Governor's Award for Excellence in Education

1998 Teacher of the Year – Central Bucks Chamber of Commerce