

**A MIXED METHOD EXPLORATION OF THE EFFICACY OF THE TEACHER  
EVALUATION SYSTEM IN SOUTHWESTERN PENNSYLVANIA**

by

**Caroline Johns**

B.S., Ohio University, 1994

M.Ed., Ohio University, 1995

Submitted to the Graduate Faculty of  
the School of Education in partial fulfillment  
of the requirements for the degree of  
Doctor of Education

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Caroline Johns

It was defended on

March 30, 2017

and approved by

Dr. James Jacob, Associate Professor, Administrative and Policy Studies

Dr. Diane Kirk, Associate Professor, Administrative and Policy Studies

Dr. Jennifer Russell, Assistant Professor, Learning Sciences and Policy

Advisor and Dissertation Chair: Dr. Cynthia Tananis, Associate Professor, Administrative and  
Policy Studies

Copyright © by Caroline Johns

2017

# **A MIXED METHOD EXPLORATION OF THE EFFICACY OF THE TEACHER EVALUATION SYSTEM IN PENNSYLVANIA**

Caroline Johns, Ed.D

University of Pittsburgh, 2017

This mixed method study was a practitioner inquiry into the teacher evaluation system that was enacted as a result of Pennsylvania House Bill 1901 (Act 82 of 2012). The study was an effort to gain an in-depth understanding of (a) factors that led to the passage of the Act 82 teacher evaluation policy, (b) what the policy was intended to do, and (c) how the policy has been translated into actual practice. Data was collected via electronic survey and semi-structured interviews. Participants were currently practicing K-12 public school principals from Beaver and Washington counties which are located in southwestern Pennsylvania. There were 27 survey participants and 6 semi-structured interview participants. Results suggest principals accurately understand the teacher evaluation tool as an accountability measure. All of them received some type of training prior to and/or during initial implementation. Effective implementation of the tool appears to be dependent on the capabilities of the principals, most of the principals are the only evaluator for their teachers and the use of student data varies by principal. Data on professional development practices was inconclusive but suggests that principals and teachers have input in district professional development offerings.

## TABLE OF CONTENTS

<b>PREFACE.....</b>	<b>XV</b>
<b>1.0 INTRODUCTION.....</b>	<b>1</b>
<b>1.1 BACKGROUND OF THE PROBLEM.....</b>	<b>1</b>
<b>1.2 PROBLEM STATEMENT .....</b>	<b>5</b>
<b>1.3 PURPOSE OF THE STUDY .....</b>	<b>6</b>
<b>1.4 RESEARCH QUESTIONS.....</b>	<b>6</b>
<b>1.5 SIGNIFICANCE OF THE STUDY .....</b>	<b>7</b>
<b>1.6 CONCEPTUAL FRAMEWORK.....</b>	<b>8</b>
<b>1.7 ASSUMPTIONS AND LIMITATIONS .....</b>	<b>9</b>
<b>1.8 DEFINITION OF TERMS .....</b>	<b>11</b>
<b>1.8.1 Coaching .....</b>	<b>11</b>
<b>1.8.2 Collective bargaining agreement (teacher contract) .....</b>	<b>11</b>
<b>1.8.3 Multiple Streams theory .....</b>	<b>12</b>
<b>1.8.4 Pennsylvania Educator Effectiveness Project (EEP) .....</b>	<b>12</b>
<b>1.8.5 Pennsylvania School Performance Profile .....</b>	<b>12</b>
<b>1.8.6 Pennsylvania Value Added Assessment System (PVAAS) .....</b>	<b>12</b>
<b>1.8.7 Plan for improvement .....</b>	<b>13</b>
<b>1.8.8 Public school.....</b>	<b>13</b>

1.8.9	Policy .....	13
1.8.10	Policy agenda .....	13
1.8.11	Professional development .....	13
1.8.12	Race to the Top (RTTT) .....	14
1.8.13	School reform.....	14
1.8.14	State assessments .....	14
1.8.15	State standards .....	14
1.8.16	Student Learning Objective (SLO).....	14
1.8.17	Tenure.....	15
2.0	REVIEW OF THE LITERATURE.....	16
2.1	POLICY DEVELOPMENT AND DEFINITION.....	17
2.2	POLICY PROCESS THEORIES .....	18
2.3	MULTIPLE STREAMS THEORY .....	18
2.3.1	Multiple Streams introduction .....	18
2.3.2	Issue definition and agenda setting .....	19
2.3.3	The participants .....	21
2.3.3.1	Institutional actors .....	22
2.3.3.2	Congress.....	22
2.3.3.3	The President and executive branches .....	23
2.3.3.4	The courts .....	25
2.3.3.5	Non-institutional actors .....	26
2.3.3.6	Media.....	26
2.3.3.7	Interest groups.....	26

2.3.3.8	Research organizations.....	28
2.4	THE THREE STREAMS.....	29
2.4.1	Problem stream.....	29
2.4.2	Policy stream .....	32
2.4.3	Political stream .....	34
2.4.4	Policy windows and coupling.....	35
2.5	EDUCATOR EFFECTIVENESS (ACT 82) .....	36
2.6	HISTORICAL BACKGROUND .....	36
2.6.1	A Nation at Risk.....	36
2.6.2	Teachers for the 21 <sup>st</sup> Century .....	38
2.6.3	Goals 2000: Educate America Act (P.L. 103-227) .....	40
2.6.4	No Child is Left Behind (P.L. 107-110).....	41
2.6.5	The politics of NCLB.....	43
2.6.6	Three Forces Changing Our Nation’s Future.....	44
2.6.7	Teacher quality and Race to the Top.....	45
2.6.8	Brookings Institution report on teacher quality .....	46
2.6.9	Race to the Top (RTTT).....	47
2.7	PENNSYLVANIA EDUCATOR EFFECTIVENESS.....	48
2.7.1	Effective evaluation of teachers.....	50
2.7.1.1	Professional development .....	53
2.7.1.2	Best practices in professional development .....	53
2.7.1.3	Models of professional development .....	55
2.8	POLICY TO PRACTICE .....	61

2.9	CONCLUSIONS .....	63
3.0	METHOD .....	65
3.1	RATIONALE FOR STUDY METHOD AND DESIGN.....	65
3.2	SETTING AND PARTICIPANTS .....	67
3.3	ETHICAL CONSIDERATIONS .....	70
3.4	DATA COLLECTION PROCEDURE.....	72
3.4.1	Study protocol .....	73
3.4.2	Pilot of the instruments .....	75
3.5	RESPONSE RATE .....	76
3.5.1	Survey distribution and response rate.....	76
3.5.2	Interview response rate .....	78
3.6	DATA ANALYSIS.....	79
3.6.1	Data management .....	79
3.6.2	Analysis method .....	80
3.7	ACCURACY AND REPRESENTATION .....	85
3.7.1	Participant issues .....	85
3.7.1.1	Interview sections.....	85
3.7.1.2	Interview scheduling.....	86
3.8	SUMMARY .....	86
4.0	DATA COLLECTION RESULTS .....	87
4.1	INTRODUCTION .....	87
4.2	PARTICIPANT INFORMATION.....	87
4.2.1	Survey participants.....	87



4.2.2	Interview participants .....	89
4.2.2.1	Interview participant background .....	90
4.2.2.2	Participant building and district information .....	90
4.3	DATA COLLECTION AND ANALYSIS OF RQ1 .....	92
4.3.1	Accountability .....	96
4.3.2	Effective evaluation tool (teacher growth) .....	100
4.3.3	Consistency .....	102
4.4	DATA COLLECTION AND ANALYSIS OF RQ2 .....	102
4.4.1	Survey analysis summary .....	103
4.4.1.1	Survey statistical analysis .....	105
4.4.1.2	Statistical testing .....	106
4.4.1.3	Testing results.....	106
4.4.2	Qualitative analysis—Interview .....	110
4.5	INTERVIEW ANALYSIS SUMMARY .....	112
4.5.1	Interview participant one.....	113
4.5.2	Interview participant two .....	116
4.5.3	Interview participant three.....	119
4.5.4	Interview participant four .....	122
4.5.5	Interview participant five .....	124
4.5.6	Interview participant six .....	127
4.6	DATA COLLECTION AND ANALYSIS OF RQ3 .....	129
4.6.1	Effective practices.....	129
4.6.2	Effective tool—Survey responses .....	131

4.6.3	Effective tool—Interview responses .....	134
4.6.4	Trained observers—Survey responses.....	139
4.6.5	Trained observers—Interview responses .....	141
4.6.6	Multiple observers—Survey responses.....	142
4.6.7	Multiple observers—Interview responses .....	143
4.6.8	Evaluation tied to student data.....	143
4.6.9	Professional development.....	145
5.0	DISCUSSION AND CONCLUSIONS .....	149
5.1	PURPOSE OF THE STUDY .....	149
5.1.1	Methods and procedures.....	149
5.1.2	Limitations .....	151
5.2	MAJOR FINDINGS AND DISCUSSION .....	152
5.2.1	Historical background.....	152
5.2.2	RQ1: Understanding principal perceptions of intended purposes .....	155
5.2.2.1	Intended purposes identified in the literature.....	156
5.2.2.2	Alignment of principal perceptions .....	158
5.2.2.3	Accountability.....	158
5.2.2.4	Effective evaluation tool .....	159
5.2.2.5	Teacher professional development .....	160
5.2.3	RQ2: Understanding how principals were prepared .....	161
5.2.3.1	Intersection of capability and instruments .....	162
5.2.3.2	Adult learning theory.....	163
5.2.3.3	Characteristics of effective professional development.....	165

5.2.3.4	Significance of continual professional learning for principals.....	167
5.2.4	RQ3: Implementation of the policy.....	168
5.2.4.1	Addressing principal capability .....	169
5.2.4.2	Professional growth opportunities.....	171
5.2.4.3	Professional dialogue .....	171
5.2.4.4	Teacher reflection .....	172
5.2.4.5	Discussions about data.....	173
5.2.4.6	Multiple observers.....	174
5.2.4.7	Software .....	175
5.3	IMPLICATIONS AND RECOMMENDATIONS.....	175
5.3.1	Implementation challenges .....	176
5.3.2	Recommendations.....	178
5.3.3	Concluding remarks .....	180
APPENDIX A	.....	182
APPENDIX B	.....	184
APPENDIX C	.....	186
APPENDIX D	.....	192
APPENDIX E	.....	198
BIBLIOGRAPHY	.....	200

## LIST OF TABLES

Table 1. <i>Data Analysis Chart</i> .....	84
Table 2. <i>Respondent Demographics</i> .....	88
Table 3. <i>Number of Years in Current Position</i> .....	88
Table 4. <i>Location of School District by County</i> .....	89
Table 5. <i>Building Scores</i> .....	92
Table 6. <i>Comparison of Intended Purposes and Principal Perceptions</i> .....	96
Table 7. <i>Responses Referring to Accountability</i> .....	98
Table 8. <i>Responses Referencing "Student Achievement"</i> .....	99
Table 9. <i>Responses Referring to "Identifying Ineffective Teachers"</i> .....	100
Table 10. <i>Responses Referencing "Effective Evaluation Tool"</i> .....	101
Table 11. <i>Responses Referring to "Consistency"</i> .....	102
Table 12. <i>Kruskal-Wallis Rank Sum Test for Intermediate Unit Training Duration by County</i>	107
Table 13. <i>Kruskal-Wallis Rank Sum Test for Intermediate Unit Training Effectiveness by County</i> .....	107
Table 14. <i>Kruskal-Wallis Rank Sum Test for Intermediate Unit Training Effectiveness by Preparedness</i> .....	109
Table 15. <i>Pairwise Comparisons for the Mean Ranks of Intermediate Unit Training Effectiveness by Levels of Preparedness</i> .....	109

Table 16. <i>Interview Questions-Instruments</i> .....	110
Table 17. <i>Interview Questions-Capacity</i> .....	110
Table 18. <i>Survey Questions-Effective Practices</i> .....	130
Table 19. <i>Interview Questions-Effective Practices</i> .....	130
Table 20. <i>Educator Effectiveness</i> .....	131
Table 21. <i>Four Domains</i> .....	132
Table 22. <i>Sample Survey Comments</i> .....	133
Table 23. <i>Professional Dialogue and Collaboration Comments</i> .....	135
Table 24. <i>Professional Reflection Comments</i> .....	137
Table 25. <i>Training and Duration</i> .....	140

## LIST OF FIGURES

Figure 1. <i>Student Enrollment</i> .....	91
---	----

## **PREFACE**

Throughout the dissertation process, I have come to appreciate the fact that I am not the sole author of this paper. Without the support of numerous people, this dissertation would never have come to fruition. First and foremost, I want to express my deep gratitude to my family. My husband went above and beyond to make sure everything ran smoothly at home so I could focus on my studies. To my son who has grown up hearing about mom going to “night school” and has rarely complained about the time I’ve dedicated to completing my degree. My parents who have provided tremendous support in so many ways. Finally, my younger sister, who became a doctor first, providing me the competitive drive I needed to persevere.

I would also like to recognize the faculty at the University of Pittsburgh who have supported me both academically and professionally. Thank you to my committee members Dr. Diane Kirk, Dr. Jennifer Russell, and Dr. James Jacob. I am truly grateful for your time and insights which added greatly to the quality of my study. Most importantly, thank you to Dr. Cynthia Tananis who has been with me on this journey since the beginning. With her guidance, I have grown as a student, professional, and person. For this I am truly grateful.

## **1.0 INTRODUCTION**

This mixed method study was a practitioner inquiry into the teacher evaluation system that was enacted as a result of Pennsylvania House Bill 1901 (Act 82 of 2012). In schools, practitioners commonly refer to the bill as "Act 82," and this is how it is referred to throughout this dissertation. The bill, in its entirety, covers a variety of areas, but the study was focused specifically on the teacher evaluation section. The study was an effort to gain an in-depth understanding of (a) factors that led to the passage of the Act 82 teacher evaluation policy, (b) what the policy was intended to do, and (c) how the policy has been translated into actual practice in the area of Pennsylvania. The practitioner inquiry approach was deemed appropriate, because the researcher is currently working in the field and has an extensive background experience with the classroom observation tool used in the new model.

### **1.1 BACKGROUND OF THE PROBLEM**

In July 2009, President Obama gave a speech introducing the Race To The Top (RTTT) competitive grant program (Jackson, 2011). The speech provided insight into the basic assumptions underlying RTTT and the subsequent Act 82 in Pennsylvania. The overarching goal was to fully equip students to compete in a global economy and fulfill their potential. President Obama stated that one of the key factors in achieving this was effective teachers. "From the



moment a student enters a school, the single most important factor in their success is the person in front of the classroom” (Jackson, 2011, p. 12). To ensure the presence of an effective teacher in every classroom, the President recommended data be used to measure effectiveness. He acknowledged that multiple measures should be used, but he emphasized the power of data to inform decision making about teachers (Jackson, 2011).

The RTTT request for proposals included a section detailing requirements focused on ensuring “great teachers” (U.S. Department of Education, 2009). Requirements included developing clear approaches for measuring student growth across all levels of the education system, designing rigorous and transparent evaluation systems, and conducting annual evaluations that included timely feedback to teachers. Information from the evaluations was required to be presented in a format intended to inform decision making regarding needed professional development and coaching. The feedback was also intended to drive (a) decisions about retaining highly effective teachers, (b) additional compensation for those who are exceptional, (c) tenure grants, and (d) removal of ineffective teachers.

In 2010, Pennsylvania Governor Rendell submitted Pennsylvania’s application for RTTT funding. In December of 2011, the U.S. Department of Education released a press notice announcing that seven states, Pennsylvania included, would each receive \$200 million in RTTT funds (U.S. Department of Education, 2011). These funds were targeted for school reforms designed to improve student achievement. This resulted in the Pennsylvania Department of Education (PDE) developing and implementing the “Educator Effectiveness Project” (Pennsylvania Department of Education, 2014), which included a teacher evaluation model based on multiple measures. In 2012, Act 82 was passed making the Educator Effectiveness model the teacher evaluation system required by law. In June of 2013, the regulations for

administering Act 82 were released. At the time of this decision, all districts were required to use the new evaluation system unless they had an existing agreement with the teacher union that detailed teacher evaluation. In those instances, districts continued to follow their standard agreement through contract expiration.

To meet the requirements of Act 82, the Pennsylvania Department of Education developed a new teacher evaluation tool comprised of multiple measures. The tool was designed to evaluate a teacher based on 50% classroom observation and 50% student performance. The classroom observation portion was based on Charlotte Danielson's four domain model (Danielson, 2007). The four domains are planning and preparation, classroom environment, instruction, and professional responsibilities. The student performance portion was based on building level data (15%), teacher specific data (15%) and elective data (20%). Building level data was determined by each individual school's Pennsylvania School Performance Profile (SPP) score. The SPP is a number comprised of multiple data points including, but not limited to, (a) student scores on the annual state exams, (b) graduation rates, (c) student attendance rates, (d) student scores on national college admissions exams, and (e) Pennsylvania Value Added Assessment System (PVAAS) growth measures. PVAAS measures the growth each student makes in one school year.

Teacher specific data was based on how each individual teacher's students performed on the state assessments, PVAAS growth of their students, how their students with special needs were progressing on their individual goals, and locally developed school rubrics. The final component was elective data determined at the district level, but approved by the Pennsylvania Department of Education. Elective measures could include district assessments, national assessments, industry certification exams, student projects, and portfolios. Elective data was

measured by teachers creating a Student Learning Objective (SLO) by which they validate and measure the elective data (Pennsylvania Department of Education, 2014).

Professional employees (i.e. tenured teachers) were required to be rated annually and temporary professional employees (i.e. non-tenured teachers) were required to be rated semi-annually. Four ratings were possible: distinguished, proficient, needs improvement, or failing. A rating of “distinguished” or “proficient” was considered satisfactory. A rating of “needs improvement” was technically considered satisfactory, but resulted in a teacher being placed on a plan for improvement. If a teacher received a second “needs improvement” within a 10 year period, it was considered an unsatisfactory rating. Both “needs improvement” ratings must also have come from the same employer while the teacher is working in the same certification area. A rating of “failing” was considered unsatisfactory. All of these requirements and conditions are in force today.

While many would agree that ensuring an effective teacher in every classroom is an admirable goal, there are difficulties with pushing education policy from the top down. President Obama set forth national goals for the country, but it is challenging to implement national goals in a country where a national system for education does not exist (McGuinn, 2006).

McGuinn (2006) refers to this as the “50/14,000/130,000 problem” in the US. There are 50 state education systems with approximately 14,000 school districts that contain approximately 130,000 schools. Each state has developed a vastly different system of educating students, and many variations exist within the districts. This results in limited capacity for state and federal government to successfully push reform to the local level. This mixed method study was designed to provide a retrospective look at policy at the federal level to determine factors that led to teacher evaluation rising in importance on the policy agenda, including intended purposes,

principals' perceptions about why teacher evaluation has risen to a high level of importance and how the evaluation system has been translated into practice.

## **1.2 PROBLEM STATEMENT**

Results of a review of the literature suggests that the classroom teacher has the greatest impact on student learning (National Commission on Teaching & America's Future, 1996); yet, ineffective teachers remain in classrooms (Gordon, Kane, & Staiger, 2006). The Pennsylvania Educator Effectiveness Project (EEP) was developed to reform how teachers are evaluated and provide the tools with which they can increase their effectiveness. In June of 2012, Pennsylvania House Bill 1901 (Act 82 of 2012), more commonly referred to as Act 82, made the evaluation system detailed in the Educator Effectiveness Project the law governing how teachers are to be evaluated.

The present study was divided into two parts. Part 1 was an exploration of factors that led to the passage of a new evaluation system and their intended purposes. This analysis was completed in the review of literature using John Kingdon's Multiple Streams theory (Kingdon, 2011). Part 2 focused on principals' experiences with the policy during the early years of its implementation. Specifically, principal's understanding of what the policy is intended to do and their experience with the implementation was explored. A review of the literature suggested that this kind of study had not been previously conducted.

### **1.3 PURPOSE OF THE STUDY**

This study was designed to be a policy analysis of the Act 82 teacher evaluation provisions through a retrospective review of policy development as well as a prospective analysis of school principal perceptions and experiences. Data collection was focused on an exploration of potential gaps in knowledge that exist regarding the policy's intended purposes and implementation. Through a review of literature, the study was begun with a retrospective look at education policy history in the United States using John Kingdon's (2011) Multiple Streams framework. The purpose of this retrospective look was to gain an understanding of why teacher evaluation became a prominent issue on the government policy agenda as well as an understanding of the intended purposes of the new evaluation system. The next part of the study focused on how the evaluation system has been translated into practice with a focus on whether or not the intended purposes have been realized in school settings. This was done by studying the perceptions, attitudes, and lived experiences of principals as they have gone through the process of implementing the evaluation system. Specifically principal's understanding of the policy and how they have translated that into practice was reviewed.

### **1.4 RESEARCH QUESTIONS**

Based on information presented in previous sections, the following research questions drove the methodology of the study:

RQ1: What are school principals' understandings of the intended purposes of Act 82?

RQ2: How have principals been prepared to implement the Act 82 Educator Effectiveness policy in their schools?

RQ3: How have principals implemented the Act 82 Educator Effectiveness policy in their buildings?

## **1.5 SIGNIFICANCE OF THE STUDY**

This mixed method study is important in two major arenas. First, there is research that suggests that teachers have the greatest impact on a child's achievement in school when compared with other factors in the school environment (National Commission on Teaching & America's Future, 1996). Second, resources in the areas of funding, human resources, and technology are required to implement the evaluation system as required by Act 82, but currently there exists a lack of research that provides a clear cost analysis. While a description is provided of what districts have done to implement Act 82, the true costs are not known at this time.

To track information, software systems had to be developed and purchased by districts to provide student data linked to individual teachers, and people had to be trained to use these new systems. Districts have had to invest human resources for determining the teacher attributions these systems track. Teacher attribution means that each student score is directly linked to the teacher(s) who provided direct instruction to the student. The process of determining teacher attributions can take a significant amount of time and human resources. Administrative time has also been invested as all administrators had to attend training workshops on how to observe and evaluate the mandates of the current model.

Early feedback indicated that the evaluation process is time consuming when completed as intended. Considering the investment of resources and the high stakes involved for children, it is important for legislators, policymakers, and practitioners to have information regarding the return on investment. This study was an attempt to provide feedback regarding the perceived benefits that have occurred as a result of the evaluation system, discover any unintended consequences, and provide recommendations for future research.

## **1.6 CONCEPTUAL FRAMEWORK**

The study commenced with a retrospective look at factors that led to the passage of the current evaluation model. In the review of literature, the Multiple Streams theory (Kingdon, 2011) was used to analyze historical events and documents leading up the passage of Act 82. The Multiple Streams theory was developed during Kingdon's study of why certain issues rise to a higher level of importance than others and subsequently gain entry on the governmental policy agenda. During the 1970s, Kingdon conducted a large scale study of development of public policy over time specifically focusing on transportation and healthcare policy at the federal level. It was during this study that the Multiple Streams theory emerged. It was theorized that at all times there are three streams running independent of each other: the problem, the policy, and the political streams. Critical time periods occur when these streams converge, or "couple," and it is during these times that an issue has the greatest chance of rising to prominence on a governmental agenda. As historical events and documents were reviewed, items were categorized into the appropriate stream. Evidence of coupling that would result in the issue of teacher evaluation rising to a level of prominence on the policy agenda was identified.

Once the historical look at Act 82 was completed and intended purposes of the policy identified, how the policy has been translated into practice was investigated. Through questions RQ1, RQ2 and RQ3, how principals have experienced implementing the policy was explored. This was done using a framework that focused on the relationship that exists when a policy moves into practice (Cohen, Moffitt, & Goldin, 2007). This relationship can be one of cooperation or conflict. Cohen et al. (2007) identified four factors that tend to determine the outcome of this relationship. Those factors are the nature of the policy, instruments, capability of the practitioners, and the environment. The nature of the policy refers to the level of clarity or ambiguity. The more clarity a policy has, the easier it can be to implement, but it may also be more restrictive in nature. The more ambiguous it is, the harder it can be to implement as intended, but may also be adaptable in a local context. Instruments refer to incentives, ideas, funding, and regulations. Capability refers to how much the policy requires a departure from current practices. The greater the departure, the greater the likelihood of increased incapability of the practitioners to implement the agenda as they grapple with new and possibly unfamiliar practices. The final one is environment, which explores factors in the environment that either support or inhibit the policy's implementation into practice.

## **1.7 ASSUMPTIONS AND LIMITATIONS**

Assumptions of this study were that all participants were licensed administrators in the state of Pennsylvania. This assumption was in place because participants were limited to people currently serving as school principals in a public school setting. It was assumed that the participants had knowledge and experience with the current evaluation model as it has been in



effect since the 2013-2014 school year. Principals working in a district where a current teacher contract exists, therefore precluding them from working under the current evaluation system, were not included in the study. The geographical location was limited to Beaver and Washington counties, both located in southwest Pennsylvania. The determination to limit the investigation to these two counties was due to several factors. Administrators in the neighboring Allegheny County, which includes the city of Pittsburgh, are more frequently asked to respond to research studies, which may be due to the preponderance of universities in the county. This has created a situation where potential participants could be less likely to respond. Beaver and Washington counties were deemed a better fit because it was thought that they were less studied, which could increase participation. As the study progressed and difficulties with response rate emerged, it was discovered that principals in these counties have also received many requests for participation in surveys. Therefore, the initial thought process was deemed incorrect.

Another factor was that fewer direct administrative relationships exist with the researcher in these counties as the researcher works in Allegheny County. It was beneficial to have no prior professional or personal experience with potential participants since a prior relationship could potentially skew results. One of the areas that was explored was supports principals have had with implementation of the policy. In Allegheny County, the researcher was already familiar with supports in place. Beaver and Washington counties provided a geographical area where the researcher had little familiarity with how the policy was implemented, yet they both remain a critical area of southwest Pennsylvania.

Limitations included researcher perceptions as a school and district administrator. I currently serve as Superintendent of Schools in a public K-12 school setting located in Allegheny County. At the start of this study, I was serving as the Assistant Superintendent in a public K-12

school setting also located in Allegheny County. As part of my professional role as Assistant Superintendent, I was in the process of planning our district's next phase of implementation of the evaluation model. This involvement led to my own perceptions regarding the positive and negative aspects of the policy, which were addressed through the coding process and subsequent analysis. Saldana (2013) provided guidance on how to monitor personal responses to data while working through the coding process. Saldana recommended that the researcher note data that causes a feeling of surprise, intrigue, and/or disturbance. Noting these reactions helps the researcher identify personal assumptions and tensions within a personal value, attitude, and belief system, minimizing bias. As I went through the coding process, I noted these reactions and provided feedback on cause.

## **1.8 DEFINITION OF TERMS**

### **1.8.1 Coaching**

A form of professional development for teachers whereby a trained professional, the coach, provides on-going feedback to the teacher on how to use effective teaching strategies in the classroom.

### **1.8.2 Collective bargaining agreement (teacher contract)**

An agreement that is reached between a local school board and teacher union detailing working conditions for teachers.

### **1.8.3 Multiple Streams theory**

A theory developed by John Kingdon in the 1970's describing how certain issues rise to prominence on the public policy agenda. Kingdon theorized that, at any given time, there are multiple streams running through government. Those streams are problems, policy, and politics. These streams run independently from each other, but there are times when they couple. This coupling creates an opportunity for an issue to rise to prominence on the policy agenda.

### **1.8.4 Pennsylvania Educator Effectiveness Project (EEP)**

Project developed with a focus on reforming how teachers are evaluated in Pennsylvania. This included a focus on using evaluation information to plan professional development opportunities.

### **1.8.5 Pennsylvania School Performance Profile**

A numerical rating assigned to a district on an annual basis to reflect how they are achieving. The number is comprised of multiple data points including, but not limited to, student scores on the annual state exams, graduation rates, student attendance rates, student scores on national college admissions exams and Pennsylvania Value Added Assessment System (PVAAS) growth measures.

### **1.8.6 Pennsylvania Value Added Assessment System (PVAAS)**

An annual measure of individual student growth on state assessments.

### **1.8.7 Plan for improvement**

A plan developed for a teacher who has been identified through the evaluation process as needing assistance. The plan details areas the teacher needs to improve upon, supports that will be provided by the district, and objectives the teacher needs to meet.

### **1.8.8 Public school**

A system of free education, grades kindergarten through 12, provided to all students that is supported by public funds.

### **1.8.9 Policy**

Outputs of a political system in the form of statutes, regulations, laws, ordinances, court decisions, executive decisions, government programs, and other constructs that result from the political system in response to a public problem.

### **1.8.10 Policy agenda**

A set of issues and problems that are viewed by public officials as worthy of being addressed.

### **1.8.11 Professional development**

Formal training provided to educators in the areas of instructional strategies, research on pedagogy, emerging technology tools, curriculum, etc.

### **1.8.12 Race to the Top (RTTT)**

A federal, competitive grant that was offered to states during the Obama Administration. The focus was on ensuring all children received an education that would enable them to graduate career and college ready.

### **1.8.13 School reform**

The process of making changes to education in an effort to improve it.

### **1.8.14 State assessments**

Standards-based, criterion referenced tests designed to track how students and schools are progressing with the state standards. In Pennsylvania, the Pennsylvania System of School Accountability (PSSA) exam is given annually in grades 3-8.

### **1.8.15 State standards**

Guidelines for what skills and content students should learn at each grade level.

### **1.8.16 Student Learning Objective (SLO)**

A measure to track student achievement on specific content. SLOs are also used as part of the teacher evaluation system in Pennsylvania

### **1.8.17 Tenure**

A status that teachers can earn after three years of probationary employment. In Pennsylvania, teachers can earn tenure after three years of successful teaching. Once a teacher is granted tenure, they have additional protections from termination or layoffs.

## **2.0 REVIEW OF THE LITERATURE**

The purpose of the review of literature is to provide background information about what policy is, how it is developed, and what constitutes the conceptual framework that serves as the foundation for the study. The first half of the discussion is an overview of policy, followed by descriptions of prominent policy process theories. The first theory presented, the Stages Model, is included because it provides an overview of the process that is described in a manner that is understandable to readers who may not have a background in policy development. Additional theories are included because the descriptions of the process capture the complexity of the policy making process. While this dissertation is an exploration of the history and implementation of Act 82 using one policy theory (Multiple Streams), it is imperative that the complexity of the policy process not be lost. The next section of the literature review contains a more detailed description of the Multiple Streams (MS) theory developed by John Kingdon. This theory serves as a conceptual framework for the present study.

The second half of the literature review contains a review of the background of teacher evaluation. This section begins with a review of historical events that led to the rise of teacher evaluation as a prominent issue and components of effective evaluation including best practices in professional development. The final section contains a summary of factors that impact how policy is translated into practice. The literature review also serves as a conceptual framework for the present study.

## **2.1 POLICY DEVELOPMENT AND DEFINITION**

The development of policy is a complex process that is utilized by political systems to address societal problems and issues. This section of the literature review begins with a description of policy development, and a definition of policy pertinent to the purpose of this study. Policy development is the process of conceptualizing problems that are brought to the government for solution. Government institutions develop alternatives (i.e. options for addressing it) and select a policy solution. Those solutions are then implemented, evaluated, and revised as needed (Sabatier, 2007). Policies are the outputs of a political system in the form of statutes, regulations, laws, ordinances, court decisions, executive decisions, government programs, and other constructs (Birkland, 2010; Kruschke & Jackson, 1987) that result from the political system in response to a public problem (Cibulka, 1995). Policy development also extends to informal practices and/or inaction on the government's behalf (Cibulka, 1995; Dye, 2001). For example, laws that exist that are no longer enforced is an example of governmental inaction.

The overall process is extremely complex and can take decades to evolve (Sabatier, 2007). Fowler (2013) provided a definition that is succinct, but also recognizes the complexities that exist. Fowler contended "Public Policy is the dynamic and value-laden process through which a political system handles a public problem. It includes a government's expressed intentions and official enactments, as well as consistent pattern of activity and inactivity" (p. 5). Fowler defined government as including elected and appointed officials at all levels of government. Examples of elected and appointed educational officials include school boards, school administrators, and classroom teachers.



## **2.2 POLICY PROCESS THEORIES**

Policy process research is a focus on the “interactions over time between public policy and its surrounding actors, events, and contexts, as well as the policy or policy outcomes” (Sabatier & Weible, 2014, p. 5). The resulting theories describe how policies are developed starting from the initial problem through policy adoption and evaluation. Some of the predominant theories include the following: Stages Model, Elite Theory, Advocacy Coalition Framework, Punctuated Equilibrium Theory, and Institutional Theory. While an in-depth analysis of these theories is beyond the scope of this dissertation, readers interested in learning more about policy processes can use this as a starting place for their own research. The next section provides an in-depth description of another predominant theory, the Multiple Streams theory developed by John Kingdon that is focused mainly on the process of issues rising to a level of prominence on the policy agenda. This theory is used as one of the conceptual frameworks for this dissertation study.

## **2.3 MULTIPLE STREAMS THEORY**

### **2.3.1 Multiple Streams introduction**

The Multiple Streams theory, proposed by John Kingdon, is one of the most cited theories of policy formation (Cohen-Vogel & McLendon, 2009). The theory focuses on the early phases of issue definition and agenda setting. Multiple Streams theory is being used as one of the conceptual frameworks for study and is therefore described in detail in the following sections.

During the 1970s, Kingdon (2011) conducted a large research project to look at the development of public policy over time. The study focused on the two areas of healthcare and transportation at the federal level. Kingdon was especially interested in the beginning stages of policy development. The central questions focused on “how the agenda is set, how the alternatives for choices are specified and why these processes work as they do” (p. 83). The study was focused on the two processes driving why some subjects become prominent on the policy agenda while other subjects do not, and why certain alternatives are seriously considered while others fade into the background. In his work, Kingdon distinguished between participants and processes. The main idea of the theory is that, at any given time, there are multiple streams running through government. Those streams are problems, policy, and politics. These streams run independently from each other, but there are times when they couple. This coupling creates an opportunity for an issue to rise to prominence on the policy agenda. Within each stream, there are multiple participants who have varying impacts on the processes. The following section provides a description of the area of policy development Kingdon focused on, which was issue definition and agenda setting. The next section describes the many participants involved at this stage of development. The final section provides descriptions of the three streams, and how they come together, or couple, facilitating the rise of an issue to a place of prominence on the government policy agenda.

### **2.3.2 Issue definition and agenda setting**

The initial stages of issue development and agenda setting are the areas of policy most educators are least familiar with since they are almost invisible to the general public (Fowler, 2013). At times, it seems like a policy has come from nowhere, and educators are left to figure out how to

implement it. Understanding agenda setting and issue definition provides educators the ability to identify events, actor actions, political factors, and other elements that may be coming together in a manner that will ultimately cause policy change.

Kingdon (2011) defined the agenda as the problems or subjects government officials are giving serious attention to at any given time. There always exist a multitude of problems facing society and government; yet, all of those problems do not get defined as issues. A problem does not make it onto a policy agenda if it is not clearly defined (Fowler, 2013; Kraft & Furlong, 2010). Even if a problem does get defined, it is not guaranteed a spot on a policy agenda. Competition to get on the agenda and stay on the agenda is fierce due to the government's limited capacity to address issues. As groups work to get their issues on the agenda, they must at the same time keep competitive issues off (Birkland, 2010). Therefore, understanding agenda setting is crucial to understanding how "groups, power and the agenda interact to set the boundaries of political policy debate" (Birkland, 2010, p. 195).

There are two key pieces to turning a demand or conflict into an issue. For the purposes of this literature review, an issue is defined as "a conflict between two or more identifiable groups over procedural or substantive matters relating to the distribution of position or resources" (Cobb & Elder, 1972, p. 82). The first is identification of a demand or conflict in such a manner that it becomes an issue the government has the ability to address. The second critical piece is that the public view it as an issue that should be addressed (Fowler, 2013). For example, during the 1950s, poverty was considered more of a personal misfortune than a public issue. This changed in the 1960s when the perception changed to one where people viewed poverty as a problem that could be addressed by the government. As a result, it became an issue,

and subsequently moved on the government agenda. This occurred despite the fact that poverty had not changed significantly during those two decades (Majone, 1989).

Cobb and Elder (1972) identified four groups for describing how issues are created: re-adjustors, exploiters, circumstantial reactors, and do-gooders. The first group, re-adjustors, use the most common method for creating issues. A political party manufactures an issue based on a perceived inequity in the distribution of positions or resources. The exploiters also manufacture an issue, but it is done for their own gain. An example of this would be a political party manufacturing an issue that is used as part of the rhetoric of their candidate's campaign. The third group, circumstantial reactors, describes an issue being created by an unanticipated event. For example, the launch of Sputnik immediately created issues in the US. The final group, the do-gooders, includes situations where people draw attention to an issue because they believe it is in the public's best interest.

### **2.3.3 The participants**

Kingdon (2011) identified a variety of groups involved in the development of policy. In the literature, these participants are commonly referred to as the policy actors. "Policy actors, or 'players' are those individual groups, both formal and informal, that seek to influence the creation and implementation of these public solutions" (Theodoulou & Cahn, 1995). The literature divides policy actors into two general groups: institutional, and non-institutional actors (Cahn, 1995; Kingdon, 2011; Theodoulou & Cahn, 1995). Kingdon used this distinction in his description, but referred to them as being either inside or outside of government. Within each of these groups there are "hundreds of actors from interest groups, governmental agencies, legislatures at different levels of government, researchers, journalists and judges" (Sabatier,

2007, p. 3). Each actor brings their own values and interests, preferences, and perceptions to the policy table. To be able to understand policy process, one must have knowledge of the goals and perceptions of the many actors across the country (Sabatier, 2007).

### **2.3.3.1 Institutional actors**

Historically, education was left almost solely to the states, but beginning in the 1960s and 70s the education policy arena become much more active (Guthrie, 1982; Kirst & Wirt, 2009). This increased activity resulted in more visible national actors. The main institutional actors at the federal level include Congress, and the President, and includes the executive bureaucracy and the courts (Theodoulou & Cahn, 1995).

### **2.3.3.2 Congress**

The US Congress is a major factor in the policymaking process due to its legislative authority, formidable publicity, longevity of members, and access to various information sources (Kingdon, 2011; Theodoulou & Cahn, 1995). Those information sources include, but are not limited to, research studies, concerns of constituents, interest group pressures, and others (Cahn, 1995; Kingdon, 2011). Congress is one of the most visible in the federal legislative branch, but there exists another group of people on Capitol Hill who also exert a strong influence. They are the congressional staff members, especially committee staff members. This group is responsible for handling many of the aspects of congressional hearings (including selecting and preparing witnesses) writing legislation, and meeting with interested parties (i.e. interest groups). They also work to gain compromise or inaction from those opposing certain legislation, prepare final reports and brief members of Congress on the bills (Sroufe, 1995). Therefore, their influence

cannot be underestimated even if they are not the ones in front of the camera making passionate pleas to pass certain pieces of legislation.

### **2.3.3.3 The President and executive branches**

The President of the US has three main avenues of influence on policy. Kingdon (2011) categorizes the three areas as follows. The first area is a set of institutional resources such as veto power and the ability to fill positions with others who also have great influence on policy. The second area of influence is in the executive branch. This branch tends towards unitary decision making with the President being the most influential factor on those decisions. The third is the President's "command of public attention, which can be converted into pressure on other governmental officials to adopt the President's agenda"(p. 25). One of the President's main avenues for impacting policy is his ability to persuade the other policy actors (Baumer & Van Horn, 2104; Cahn, 1995; Neustadt, 1990). One only needs to look as far as some of presidential speeches to find evidence of this. Because any speech given by the President draws media attention, they are able to leverage this attention to an issue and keep attention on it (Baumer & Van Horn, 2014; Fowler, 2013). The following quotes from speeches show the increasingly active role of President in policy agenda setting:

President George W. Bush: I'd like to be the education president...our schools are absolutely not as good as they must be...and to achieve quality results, we must set and enforce standards, provide incentives, and permit the freedom and flexibility on the local level to experiment with new ideas.

President Bill Clinton: All Americans should be able to read on their own by third grade, every single one of them.

President Barack Obama: ...if you set and enforce rigorous and challenging standards and assessments; if you put outstanding teachers at the front of the classroom; if you turn around failing schools-your state can win a Race to the Top grant that will not only help students outcompete workers around the world , but let them fulfill their God-given potential.

When it comes to policy actors, the President has the ability to put himself center stage and command attention. Meanwhile, backstage there are a number of actors in the executive bureaucracy who also have a large influence on policy development. The results from Kingdon's study suggested that the presidential staff is highly influential in agenda setting. Currently, this group consists of 1,600 higher level positions that are filled by political appointment. Within that group, the closest to the President are the agency executives who are housed in the Executive Office (Fesler, 1983). The President is able to use all of these appointees to push his policy agenda in a top-down approach (Baumer & Van Horn, 2104).

The top-down approach to getting a policy agenda enacted is more complicated than it appears. It is incorrect to interpret this model as a straight top-down approach with the President setting the agenda and his political appointees faithfully enacting it. In reality, this process is much more complex and in some cases looks more like a tug-of-war as opposed to a linear approach. For example, the President appoints heads of departments and bureaus, but within those organizations there exists career civil servants who typically have more expertise due to their longevity in this area (Fesler, 1983). Political appointees arrive with the intention of remaining loyal to the Presidential agenda, but this can be difficult once they are immersed in the organization. There are situations where Presidential appointees "turn native" as they become more influenced by the priorities of the organizations they oversee as opposed to focusing on the President's priorities.

Another factor that adds to the complexity is the career bureaucrat who exists in these organizations. They exert a strong influence due to their longevity and expertise. They also have power due to their strong relationship with legislative committees and interest groups. This triad is referred to as the “iron triangle” in the policy literature (Fesler, 1983). Despite these various influences, staff comprising the executive bureaucracy still remain a powerful force in moving forward the President’s political agenda.

#### **2.3.3.4 The courts**

The courts can have considerable influence on existing policies even though they are not necessarily involved with the development of them (Baumer & Van Horn, 2104). When developing education policy, the main actors wrestle with questions that revolve around “should we,” whereas the courts answer the questions “may we” and “must we” (Mead, 2009). Therefore, policy and law as concepts are “inextricably bound together in the cycle of activity that determines what public education is and should be in the US” (Mead, 2009, p. 294). Courts interpret the laws which “can alter existing legal rules, and they can reshape or even overturn policies made by the legislature and executive branch” (Theodoulou & Cahn, 1995, p. 281). An example of the courts playing the role of policy actor is evident in the *Brown v. the Board of Education* (1954) case where the courts determined that separate was not equal when educating children from different racial backgrounds. This decision put states on notice that the courts would not hesitate to scrutinize educational policy that did not align with the protections provided by the U.S. Constitution (Mead, 2009).



### **2.3.3.5 Non-institutional actors**

Outside of the formal institutions, there exist groups of people who also impact policy. This section provides several examples of those groups and the role they play in policy development. While they may not be part of a formal institution, their influence cannot be underestimated. They often interact with the institutional policy actors influencing them in a multitude of ways.

### **2.3.3.6 Media**

The media has a powerful influence on the issues people pay attention to that impacts the policy agenda setting process (Baumer & Van Horn, 2014; Cahn, 1995; Iyengar & Kinder, 1987; Stromberg & Snyder, 2008). The media tends to influence how people perceive reality because of how people process information. This makes them especially vulnerable to the influence of media (Graber, 1988). As people come into contact with information, they tend to pare it down. They then interpret the information in a manner that fits with their own background information since news stories generally do not provide background. They also tend to attach more weight to issues being covered in the media. Ultimately, their voting behavior is impacted. As a result, policymakers are motivated to pay attention to issues being covered by the media (Stromberg & Snyder, 2008). The reverse is also true. Issues that are not covered by the media tend to be ignored (Baumer & Van Horn, 2014).

### **2.3.3.7 Interest groups**

Interest groups are groups of people who unite to promote their political goals and outcomes (Birkland, 2011). Since the 1960s, interest groups have grown remarkably. Between the 1960s and 1990s growth of interest groups rose by 300 percent (Walker, 1991) and the number of groups doubled between 2000 and 2005 (Kernell, Jacobson, & Kousser, 2011). There are

different types of interest groups including, but not limited to, private economic interest groups, citizen groups, and groups hired by state and local officials to represent them in Washington (Baumer & Van Horn, 2014). Interest groups work to lobby elected officials and government administrators to pass policies that either benefit them or stop ones that would have an adverse effect on their group (Baumer & Van Horn, 2014). Some of the strategies used by interest groups are testifying at Congressional hearings, observing committee deliberations, using public campaign strategies to get voters to support their cause, and contributing to political campaigns in order to gain favor for their causes (Baumer & Van Horn, 2014). As a result, interest groups can have considerable influence on the institutional actors who play a large role in policymaking.

In the education realm, interest groups have grown increasingly specialized in terms of their interests and less focused on the common goals of public school. Groups such as the Parent Teacher Association (PTA) have lost ground to specialized groups such as those representing interests in the education of handicapped children, Christian organizations, parents of limited-English-speaking children, and others (Kirst & Wirt, 2009). In their book, *The Political Dynamics of American Education*, Kirst and Wirt (2009) divided education interest groups into three broad categories. The first is professional educators with national examples being the National Teachers Association (NEA) and the American Federation of Teachers (AFT). The second group is those who see education as a means to other ends. These are groups who wish to use school “to serve other values, such as reducing taxes, protecting moral or patriotic values” (p. 81). For example, the Business Roundtable is comprised of the CEOs of some of the nation's largest corporations. This group previously committed to supporting No Child Left Behind because they viewed it as a means to ensuring economic competitiveness. The final category is policy issue networks that are composed of various interest groups supporting one issue. For

example, in the 1990s groups including academics, think tanks, foundations, and the AFT joined together to support the development of academic standards and testing. Despite the increase in groups with a specialized interest, teacher unions still remain by far the most powerful of the education interest groups (Fowler, 2013).

#### **2.3.3.8 Research organizations**

Research organizations, more commonly referred to as think tanks, operate less in the public eye as compared to the media and interest groups, but they are no less influential (Anderson, 2011). Think tanks are independent research organizations often dependent on foundations, corporations, and individuals for financial support (Birkland, 2010, Fowler, 2013). They also exist in a number of universities (Anderson, 2011). Think tanks exert considerable influence because “In the process of deciding what problems to study, selecting sources of information about them, and summarizing the results of their research, the staffs of think tanks quietly determine which social problems will be considered public policy issues and how those issues will be conceptualized, both by policy makers and by the general public” (Fowler, 2013, p. 139). Their research provides policymakers with information on policy issues, alternatives and proposals for addressing the issues as well as evaluating the consequences of public policies (Anderson, 2011). Examples of notable think tanks are the Brookings Institution, the Cato Institute, and RAND Corporation (Anderson, 2011; Birkland, 2010).

## **2.4 THE THREE STREAMS**

This section describes the three streams Kingdon identified. They normally run independent of each other, but there are times when they join, or couple, together. It is during these critical periods that an issue can rise to prominence on the policy agenda. Below is a detailed description of each policy stream followed by a description of what occurs when the streams couple resulting in the opening of a policy window.

### **2.4.1 Problem stream**

In the problem stream, Kingdon (2011) focused on why some problems capture public officials' attention while others remain unaddressed. A problem is defined as a condition that people feel should be addressed. There always exist a multitude of problems in society, but some rise to a higher level of importance. Kingdon proposed three avenues for problems capturing attention: monitoring of indicators, focusing events and feedback.

Indicators can alert government officials to a problem, especially if there is a change in an indicator. Officials are constantly monitoring a number of indicators such as federal expenditures and budgetary impacts. A pattern of increasing costs can be an indicator of a problem (Kingdon, 2011). At the state level, officials and educators are constantly monitoring indicators of student achievement such as standardized testing scores. A decrease in scores across the state would indicate a potential problem. It would not, however, indicate the nature of the problem or the cause.

Focusing events can also draw attention to a problem. A focusing event can be a crisis or disaster or a powerful symbol that catches on (Kingdon, 2011). Cobb and Elder (1972) referred

to these focus events as a triggering events. In their work, they identified two major groups of triggering mechanisms (unforeseen events): internal, and external. Examples of internal events are natural catastrophe, unanticipated human event, technological changes, imbalance in distribution of resources and ecological change. Examples of external events include an act of war or violence against the US, innovations in weapons technology, international conflict not involving the US, and changing international allegiances.

Political symbols, including language, can also be used to draw attention to problems. Policy actors can strategically utilize symbols to bring attention to issues that are important to them. A symbol can be a word, phrase, gesture, event, person, place, or thing (Elder & Cobb, 1983). Consider the word “Watergate.” The word refers to an apartment complex located in Washington, D.C., but many people immediately associate it with the 1972 break-in scandal at the Democratic Headquarters. This common association demonstrates how “An object becomes a symbol when people endow it with meaning, value or significance” (Elder & Cobb, 1983, p. 29).

Symbolic language is a specific tool that policy actors can use to draw public attention to problems or issues. It has the ability to shape the public’s perception of events. Evidence of this can be found throughout history. “From the beginnings of recorded history to the present day, governments have won the support of large numbers of their citizens for policies that were based upon delusions: beliefs in witches, in nonexistent internal and external enemies....etc.” (Edelman, 1977, p. 3). Language has the power to create and form political realities for citizens. The public is especially vulnerable when there is high anxiety. By focusing on a threat and then providing reassurance, political leaders are able to gain followers (Edelman, 1977). For example, political actors can bring attention to a potential issue by describing it as being a threat

to “national security.” Using symbols, groups can skillfully draw attention to problems, define them as they wish and develop proposed solutions.

Feedback is the third mechanism for drawing attention to a problem. Feedback typically comes from programs that are not working as planned, implementation of legislation that does not occur as intended, problems that arise as a result of a new program being enacted or unanticipated consequences. This feedback can come back to officials through systemic monitoring and evaluation or through informal routes such as citizen’s complaints to legislators (Kingdon, 2011). An example of feedback can be found in what occurred after the 2002 passage of No Child Left Behind (NCLB). One of the aspirations of NCLB was to ensure that all students made academic progress. This included an emphasis on racial and ethnic subgroups. States were required to set a minimum proficiency rate on standardized exams. School with subgroups failing to meet those minimum proficiency rates were deemed to not make Adequate Yearly Progress. Kane and Staiger (2003) found that the requirements of NCLB, as initially implemented, did not make a significant difference in closing the achievement gap. At the same time, a large number of schools with more diverse populations were designated as failing. Kane and Staiger (2003) concluded that the NCLB subgroup rules were actually counterproductive to student achievement and resulted in more sanctions and fewer resources, especially in diverse schools.

Kingdon (2011) makes a special note about budgets in his description of the problem stream. The budget can act as a constraint or a promoter. The budget acts as a constraint when the possible solutions to a problem are costly. On the other hand, it can act as a promoter when a solution is available that helps contain costs. For example, if costs are rising in a certain area

and possible solutions exist to contain those costs then that problem has more likelihood of rising on the agenda. More frequently, the budget functions as a constraint.

### **2.4.2 Policy stream**

The policy stream is the second stream, and it runs independent of the problem stream most of the time. Kingdon (2011) identified the actors in this stream as the research organizations, congressional staffers, academics, interest analysts, people involved with planning and evaluation, and people involved with budgetary issues. Actors from these various organizations who have an interest in a common policy area combine to create a community of specialists. Actors within a community are constantly interacting with each other, exchanging ideas, creating new ones and modifying existing ones. Kingdon noted that these communities tend to be either cohesive or fragmented. Cohesive communities often develop common outlooks and ways of thinking. On the other hand, fragmented communities tend to experience greater agenda instability. The more cohesive a policy community is the more likely they are to generate viable alternatives.

The main focus of policy communities is to consider a large number of policy alternatives and narrow them down to serious proposals (Kingdon, 2011). Alternatives are possible solutions to problems. Kingdon uses the term “policy primeval soup” to describe this process. A multitude of ideas float around in a policy community being discussed, revised, and joined together in new ways. Through this “softening up” process some ideas float to the top as viable alternatives as other others sink to the bottom and potentially disappear. Alternatives that survive typically have five common characteristics. These alternatives are technically feasible,

aligned with the values of the policy community, fit within budget constraints, will be acceptable to the general public, and will be supported by elected officials.

Alternatives that are technically feasible are ones that have implementation details worked out. Policy members have discussed if the implementation plan meets the intended goal of the alternative. The alternatives that emerge from a community also are aligned with the values of the specialists developing them. Examples of common value themes are the role of government versus private sector or equity of distribution of resources. The third characteristic relates to budgetary constraints. If an alternative is outrageously expensive it is unlikely to have support from elected officials and the public despite technical feasibility and value alignment. Therefore, an alternative must have acceptable costs. The final two characteristics are tied together. An alternative needs to be acceptable by the general public and elected officials if it is going to survive the primeval soup. Policy specialists often view the elected officials' acceptance of an alternative as indicative that it will also be accepted by the public (Kingdon, 2011).

In this stream, there also exists policy entrepreneurs. These entrepreneurs come from a variety of backgrounds but have one common characteristic. They advocate strongly for a particular policy solution. They are "willing to invest their resources-time, energy, reputation, and sometimes money-in the hope of a future return" (Kingdon, 2011, p. 122). There are several reasons why policy entrepreneurs advocate including promotion of personal interests, desire to promote their values by shaping public policy and/or they simply enjoy participating in the process (Kingdon, 2011). Policy entrepreneurs become especially important when the three streams converge as is described in a later section.



### **2.4.3 Political stream**

The political stream is the final stream. While the other two streams focus more on problems and their solutions, this stream impacts which issues make it onto the political agenda. The components that make up this stream are public mood, pressure from organized political groups (i.e. interest groups), election results, shifts in the ideology of the majority congress and changes in administration (Kingdon, 2011). Kingdon describes national mood as the idea that at any given time there are a large number of people thinking along the same lines, the mood changes at times and it has an impact on policy agendas and outcomes. He also found during his study that government actors speak about the national mood indicating that they have a good sense of it. An example of national mood is if there is a large financial crisis occurring in the country, the national mood may indicate a desire for strong government intervention. Conversely, there are times when the national mood leans towards less government action. Government actors keep a pulse on this mood and use it to determine when to push certain items on their policy agenda and when to keep ones off.

The second component of the political stream is pressure from organized political groups. This includes pressure from interest groups, political mobilization, and behavior of political elites (Kingdon, 2011). Government officials pay attention to what these groups are advocating for and what they oppose. A policy proposal has a higher chance of success if a number of groups are voicing support for it. If there is a mixed reaction, the proposal may still be successful, but with an understanding that there may be a political price to be paid. If there is unified opposition, the advocate(s) for the proposal will likely retreat and the proposal will disappear. Advocates for proposals are constantly weighing the benefits of getting their proposals on the agenda versus the political costs of doing so (Kingdon, 2011).

The final components of election results, shifts in party majority, and administration changes are closely related. Newly elected officials quickly make their agendas known. Incumbents publicize changes in priorities and new agenda items (Kingdon, 2011). For example, when a new President is elected one of the first things he does is talk about his policy priorities. This has a great impact on which proposals have the potential to move forward and which ones fade into the background. Of course, the President's ability to move his agenda forward is affected by the political composition of Congress. Unlike the policy stream, building consensus in the political stream is often more about bargaining as opposed to softening.

#### **2.4.4 Policy windows and coupling**

Most of the time the three streams run independent of each other, but there are periods of time when they join or couple together. Kingdon refers to these important periods as policy windows. A policy window is a short, but critical time when three factors align. An example of coupling is when a problem exists (problem stream) for which there is a ready solution (policy stream) and this occurs during a favorable political climate (political stream). It is possible for an item to make it onto the agenda when only the problem and policy streams couple, but it is more likely to rise to prominence when all three streams align. Once the window closes, it is very difficult to get an item on the agenda, and it may be a very long time before the window opens again (Kingdon, 2011). Policy entrepreneurs were described in an earlier section, and it is during this coupling process that they become significant. They facilitate the coupling of streams as they “hook solutions to problems, proposals to political momentum, and political events to policy problems” (p. 182). Without the advocacy of these actors, the coupling may not occur.

## **2.5 EDUCATOR EFFECTIVENESS (ACT 82)**

Historically, education policy was a state function, but as this section of the literature review shows, the federal government has increased its influence on the states. This section begins with a review of historical events and documents that show an increasing focus on teacher effectiveness, and how events at the federal level influenced Pennsylvania's development of Educator Effectiveness. The next part reviews the literature on effective evaluation, including the literature supporting the components that comprise effective evaluation.

## **2.6 HISTORICAL BACKGROUND**

### **2.6.1 A Nation at Risk**

Beginning in the 1980s, a reform in education movement gained momentum beginning with the release of the 1983 A Nation at Risk: The Imperative for Educational Reform (ANAR) report and culminating in the passage of the No Child Left Behind Act (NCLB) of 2002 which dramatically increased the involvement of federal government in education (McGuinn, 2006; Peterson & West, 2003). In 1981, the Secretary of Education created the National Commission on Excellence in Education with the directive to develop a report on the quality of education in America for the people of America (National Commission on Excellent in Education, 1983). This signaled a political shift for then President Ronald Reagan who at the beginning of his presidency was moving towards decreased federal involvement in education with responsibility

shifting completely to the state and local levels. This course was reversed as the ANAR was received with great support (Peterson & West, 2003).

The ANAR began with a letter to Secretary Bell stating that the purpose of the report was to “define problems afflicting American education and to provide solutions, not search for scapegoats” (p. 2). The report began by describing a rising level of mediocrity in the schools that citizens needed to be concerned about due to the increase in international competition and the higher level of skills needed by American workers so the country could remain economically competitive. One section details the “indicators of risk,” which includes declining test scores, decreased rankings compared to other countries, increasing rates of illiteracy, underdevelopment of gifted students, lack of higher order thinking skills, increasing need for remedial classes at the college level, and military and business leaders complaining that students do not come with the basic skills needed to be successful. The report was a summary of the issues with a strong statement indicating that for the first time in the nation’s history the education skills of the current generation would be seriously lacking compared to their parent’s generation.

The impact was felt quickly. In the problem stream, education reform issues took a prominent position in the public sphere (Hunt & Staton, 1996) and the amount of media coverage on education issues increased significantly (Dominick, 1984). Scholarly debates exist about whether or not ANAR had an impact on education, but what is interesting is the impact it had on public discourse focused on educational reform. In this case, a commission was formed and in a strongly worded report indicated that a crisis existed. Within the report numerous rhetorical ironies existed setting the stage for on-going dialogue and debate. For example, some argued that the language was alarmist yet the recommendations were cautious and conservative (Hunt & Staton, 1996). In the political stream, another impact ANAR had was “raising educational

issues higher on state political agendas” (Peterson & West, p. 6). State governors discovered that the new focus on education was good for business. They were able to balance their liberal and conservative supporters by increasing educational funding while simultaneously increasing the focus on accountability (Peterson & West, 2003).

## **2.6.2 Teachers for the 21<sup>st</sup> Century**

Several years after the release of ANAR the focus went directly to teachers. As concerns about the state of education continued, the Carnegie Forum on Education and the Economy was established to bring attention to the connection between economic growth and the abilities of the country’s citizens. The forum created the Task Force on Teaching as a Profession to examine the teaching profession. The report developed by the task force was *A Nation Prepared: Teachers for the 21st Century*. The opening paragraph reads like a doomsday scenario for the future of the US:

America’s ability to compete in world markets is eroding. The productivity growth of our competitors outdistances our own. The capacity of our economy to provide a high standard of living for all our people is increasingly in doubt. As jobs requiring little skill are automated or go offshore, and demand increases for the highly skilled, the pool of educated and skilled people grows smaller and the backwater of the unemployable rises. Large numbers of American children are in limbo-ignorant of the past and unprepared for the future. Many are dropping out-not just out of school but out of productive society. (Task Force on Teaching as a Profession, 1986, p. 1)

The authors stated that in the past the people always turned to education when faced with a crisis, and this situation was no different. In this crisis, two truths were essential to

understanding how to successfully address the issues. The first was based on the need for more rigorous academic standards and the second focused on teachers. To address this crisis, there was a need to create “a profession equal to the task—a profession of well-educated teachers prepared to assume new powers and responsibilities to redesign schools for the future” (Task Force on Teaching as a Profession, 1986, p. 2).

The report laid out a utopian plan designed to prepare teachers prior to entry into the profession and then remove all barriers that might impede their work with students. The plan was structured around eight major elements. Some of the elements focused on attracting high quality candidates to the field by increasing salaries and career opportunities. This included an increased focus on preparing and attracting minority candidates. College students pursuing education careers would be required to earn a Bachelor’s degree in the arts and sciences prior to studying teaching and a new Master in Teaching degree would be developed. Within the schools, restructuring would enable teachers to make the big decisions about how to meet state and local goals, and lead teachers would assist with the redesigning of schools and upholding their colleagues to high standards. Teachers would be provided with incentives for performance and all of the resources needed in terms of technology, services, and support staff would be provided. Finally, a National Board for Professional Teaching Standards would be created to develop high standards for teachers and recognize those who met them. While the report began with gloom and doom, it quickly took a turn to the optimistic with the focus on providing a higher level professional environment for teachers so they could get on with the heroic effort of saving the nation.

### **2.6.3 Goals 2000: Educate America Act (P.L. 103-227)**

In 1994, Congress passed Public Law 103-227 more commonly known as Goals 2000: Educate America Act. This law demonstrated the continued presence of the federal government in deciding education policy direction, but in this case there was an increased focus at the state level. The main tenet of Goals 2000 was to create “new national structures to guide states toward a national strategy” (Schwartz & Robinson, 2000, p. 174). The ANAR report made education a national security concern. Goals 2000 indicated that the problem would be addressed at the state level with federal guidance (Schwartz & Robinson, 2000). The legislation provided grants for “implementation of State systemic reform under which States will develop and implement reform plans, State content and performance standards, opportunity-to-learn standards or strategies, and assessments” (Stedman, 1994, p. 1).

The foundation of the law can be traced back to President Bush’s administration which worked collaboratively with the National Governors’ Association to develop six national goals (Schwartz & Robinson, 2000). Goals 2000 turned these goals into legislation and also added two additional goals focusing on professional development for teachers and parental involvement (National State Association of Special Education Directors, 1994; Stedman, 1994). The goals addressed school readiness, increased high school graduation rate, demonstrated competency at grades 4, 8 and 12, US students scoring first in the world in math and science, literacy skills to compete in a global economy, safe school environment, increased parental involvement and development of the teaching force (Stedman, 1994). Overall, this legislation introduced state standards, standardized assessments and accountability for all students but for the purposes of this review what is notable is the focus on the teachers. Goal number 7 stated that “The nation’s teaching force will have access to programs for the continued improvement of their professional

skills and the opportunity to acquire knowledge and skills needed to instruct and prepare all American students for the next century” (Stedman, 1994, p. 2). Within this goal there were four component parts. The first was pre-service and in-service teachers have access to professional development to learn how to effectively teach an increasingly diverse population. The second was opportunities to gain the knowledge to teach challenging subject matter and use new forms of methods, assessments, and technology. The third was focused on recruiting and retaining highly talented teachers and the final was the development of partnerships to provide professional development (National State Association of Special Education Directors, 1994). What this legislation recognized was that having a highly qualified teacher in the classroom mattered.

#### **2.6.4 No Child is Left Behind (P.L. 107-110)**

In 2001, Public Law 107-110 titled, *To Close the Achievement Gap with Accountability, Flexibility, and Choice, So No Child is Left Behind* went into effect. This law was an amended version of the Elementary and Secondary Act of 1965 and became more commonly known as No Child Left Behind (NCLB). Accountability was one of the central themes of the law (Hess, 2003; Linn, Baker, & Betebenner, 2002; Peterson & West, 2003; Rudalevige, 2003), but it was not a new theme in the political realm. Accountability requirements were first imposed under the Clinton administration in the passage of a law in 1994, but the law was vague and never truly enforced. What it did accomplish was creation of bipartisan support for school accountability which was visible in the 2000 presidential campaign (Peterson & West, 2003) and laid the path for the eventual passage of NCLB.



Prior to the passage of NCLB, states had been developing their own content standards leading to inconsistencies (Hamilton, Stecher, Marsh, McCombs, Robyn, Russell, Naftel, & Barney, 2007; Linn et al., 2002). The new law now required all states to demonstrate that their students were making increases in achieving proficiency on these standards. Specifically, states had to implement testing in grades 3-8 in math and reading and demonstrate consistent gains each year known as achieving “adequate yearly progress” culminating in 100% proficiency within 12 years (Hamilton et al., 2007; Linn et al., 2002; Rudalevige, 2003). If a school or district failed to make AYP, they were required to employ scientifically based researched instructional strategies and/or programs (Linn et al., 2002).

After two consecutive years of not making AYP, a district would be required to develop a formal plan of improvement and offer parents the choice of having their child transfer another school. At 3 years of inadequate progress, districts would be required to provide supplemental services and in subsequent years the consequences increased to potential staffing changes and government takeover (Hamilton et al., 2007; Rudalevige, 2003).

Another new requirement in this law was the focus on closing the achievement gap. Districts were now required to report testing results based on subgroups including low socioeconomic status, race-ethnicity, English language proficiency and disability (Linn et al., 2002; Rudalevige, 2003). NCLB signaled a shift in who would determine the direction of school direction and policy. In the 1960s, federal government increased its presence in education by providing increasing levels of funding. During subsequent decades, more direction on how schools should operate was provided by the federal government, but it was with the passage of NCLB that there was significant shift in regards to the federal government influencing state level policy.

### **2.6.5 The politics of NCLB**

The overall intent of the law was to create a “school based accountability” system (Hamilton et al., 2007). There were several factors at play that created conditions whereby this law was able to pass in Congress. These factors provided a situation where Democrats and Republicans could support one issue (accountability) despite the fact that they supported it for different reasons. There was a common belief that despite billions of dollars being poured into education student achievement had not made any substantial gains. The accountability piece of NCLB provided Democrats a way to continue to provide resources because there was attached expectations for reform, and Republicans were able to justify resources because of the new system of accountability (Rudalevige, 2003).

The two political parties may have come together to back accountability, but there were also other groups who formed in opposition to the changes, especially as unintended, negative consequence became evident. There are four groups who typically emerge to resist high stakes reform: educators, ethnic and socioeconomic communities who are negatively impacted, high achieving school districts and supports of area of curriculum that become marginalized due to the increased focus on tested content area (Hess, 2003). As NCLB became fully implemented there were positive results revolving around the focus on student achievement, but there were also negative consequences such as decline in staff morale and a narrowing of the curriculum due to the increased focus on tested content areas (Hamilton et al., 2007). Once this kind of tug-of-war begins and politicians get resistance the common result is “In the face of heated opposition, proponents often agree to a series of compromises on program design and implementation, eventually undercutting the coercive promise implied by high-stakes testing” (Hess, 2003, p. 60). The initial lofty ideas eventually get compromised to something much less than what was

initially envisioned. On the other hand, there is a balance as policy starts to become practice and becomes institutionalized resulting in it being an accepted part of how educators do school.

### **2.6.6 Three Forces Changing Our Nation's Future**

In 2007, a dire forecast was issued for the US. According to researchers at the Educational Testing Service (ETS) the country was in the midst of a perfect storm with a grim outlook ("America's Perfect Storm," para. 1). A report issued by ETS entitled *America's Perfect Storm: Three Forces Changing Our Nation's Future* (Braun, Kirsch, Sum, & Yamamoto, 2007) outlined three forces that were converging putting the country on a perilous course towards disaster. The forces outlined in the report are divergent skill distributions, the changing economy, and demographic trends. The first force, divergent skill distribution, refers to the US high school graduation rate as compared to other Organization for Economic Co-operation and Development (OECD) countries, the growing disparity between US minority populations, especially in the areas in math and reading, and for the general population as a whole the lack of math and reading literacy skills needed for an increasingly competitive work environment.

The second force, changing economy, refers to the changes that were occurring due to technological innovations and globalization. The third force, demographic trends, refers to a population that the report said was growing older and more diverse. These three forces interacting was considered cause for great concern because the prediction was that over the next 25 years more highly educated Americans would be leaving the workforce at a time when jobs increasingly required high level skills. According to the report, those entering the workforce would do so less educated and therefore less prepared to meet the demands of the changing economy. This placed the nation in perilous risk considering the rising competitiveness of other

countries. Despite the dire warnings and threatening language, the report ends on a hopeful note as captured in the following call to action:

Given the forces described in this report, a looming question is whether we will continue to grow apart or, as a nation, we will invest in policies that will help us to grow together. We strongly believe the latter is the better course of action over the long term. While new policies focusing only on education and skills will not solve all the challenges associated with existing inequalities, if our society's overall levels of learning and skills are not increased and the existing gaps are not narrowed, there is little chance that economic opportunities will improve among key segments of our population. The call to action was sound educational policies that would increase the level of learning. (Braun et al., 2007, p. 26)

### **2.6.7 Teacher quality and Race to the Top**

During the 2000s, the streams came together opening a policy window for Race to the Top. The problem stream described in the previous section predicted a dire future for the US due to the lack of growth of a highly qualified workforce. Without this the US would lose its ability to compete in a global economy as other countries continued to build an increasingly educated workforce. Something needed to change in the education system in order to better prepare future workers and ensure our competitiveness in a global economy.

### **2.6.8 Brookings Institution report on teacher quality**

In the policy stream, the Brookings Institution published a report in 2006 titled *Identifying Effective Teachers Using Performance on the Job* authored by Robert Gordon, Thomas Kane, and Douglas Staiger. The report focused on increasing teacher quality through five recommended actions. The first recommendation was to make it easier for people to enter into the teaching profession through a non-traditional path. The authors made the assertion that there is not a correlation between completing a traditional teacher preparation program and being highly effective in the classroom. Therefore, this provided an alternate pathway for people with an undergraduate degree and specific subject knowledge the ability to enter the field.

The second recommendation was to make it more difficult for the least effective teachers to earn tenure. Included in this recommendation is linking student data to specific teachers in order to determine level of effectiveness. This would be done by using student data to determine “value added” by individual teachers. There were several suggestions on how this could be implemented, but the basic premise was that the least effective teachers would be evaluated out. A principal could choose to maintain a teacher but would have to provide publicly a rationale for why that teacher was being retained. The authors made the claim that the economic value of doing this would be “enormous” due to the increased student achievement gains.

The third recommendation was to provide financial incentives to highly effective teachers willing to work in areas with a high proportion of low income students. It would not be sufficient enough for a teacher to be willing to work in these areas. The teacher would need to have a proven track record of effectiveness based on student data.

The fourth recommendation focused on teacher evaluation, specifically using multiple measures. The authors recommended that student achievement be a substantial factor in

evaluations, but they also acknowledged that no single measure is perfect. Therefore, they recommended using multiple measures such as principal observations, review of student work, parent feedback, etc. One notable piece of this recommendation was the authors' stated that this kind of rigorous evaluation system would never be successful without it being perceived as fair by teachers. To ensure this the authors recommended that teachers fully participate in the development of these measures.

The final recommendation was to develop data systems to link student performance with individual teachers in order to measure effectiveness. NCLB had required states to test in grades 3 through 8 but very few states actually linked those results to teachers. The authors recommended that the federal government expand their support to help states put these systems in place.

### **2.6.9 Race to the Top (RTTT)**

In the political stream, a new President was elected in November of 2008. President Obama maintained close ties to a political action group called the Democrats for Education Reform (DER). Prior to his election, Obama announced to members of this group that if the Democratic Party had a position on education he could not figure out what it was and the party needed to wake up and figure it out (Spring, 2014). Subsequently, the DER issued a series of briefs with recommendations based on the Brookings report. These briefs were sent to U.S. Secretary of Education Arne Duncan in 2009. The ideas covered in the briefs were incorporated into Race to the Top (Spring, 2014). In a speech given by President Obama during 2009, he outlined the main tenets of RTTT.

The race starts today. I'm issuing a challenge to our nation's governors, to school boards, and principals and teachers, to businesses and non-for-profits, to parents and students: if you set and enforce rigorous and challenging standards and assessments; if you put outstanding teachers at the front of the classroom; if you turn around failing schools-your state can win a Race to the Top grant that will not only help students outcompete workers around the world, but let them fulfill their God-given potential. (Jackson, 2011)

RTTT was designed as a competitive grant process in an attempt to circumnavigate the issues occurring with the reauthorization of NCLB due to Congressional gridlock. One of the notable features of this policy is it came directly from the executive office. The President, as a policy actor, always maintained great political influence, but President Obama's actions took this to a new level (McGuinn, 2006).

## **2.7 PENNSYLVANIA EDUCATOR EFFECTIVENESS**

In 2010, Pennsylvania Governor Rendell submitted Pennsylvania's application for RTTT funding. In December of 2011, the U.S. Department of Education released a press notice announcing that seven states, Pennsylvania included, would each receive \$200 million in RTTT funds (U.S. Department of Education, 2011). These funds were targeted towards school reforms designed to improve student achievement. This resulted in the Pennsylvania Department of Education (PDE) developing and implementing the Educator Effectiveness Project (Pennsylvania Department of Education, 2014) that included a teacher evaluation model based on multiple measures. In 2012, Act 82 was passed making the Educator Effectiveness model the new teacher evaluation system as required by law. In June of 2013, the regulations for

administering Act 82 were released. At this time, all districts were required to use the new evaluation system unless they had an existing agreement with the teacher union that detailed teacher evaluation. In those instances, districts continued to follow their current agreement through contract expiration.

To meet the requirements of Act 82, the Pennsylvania Department of Education developed a new teacher evaluation tool comprised of multiple measures. With this tool a teacher's evaluation is based on 50% classroom observation and 50% student performance. The classroom observation portion is based on Charlotte Danielson's *Framework for Teaching*, which divides effective practice into four domains (Danielson, 2007). The four domains are planning and preparation, classroom environment, instruction, and professional responsibilities. The student performance portion is based on building level data (15%), teacher specific data (15%) and elective data (20%). Building level data is determined by each individual school's Pennsylvania School Performance Profile (SPP) score. The SPP is a number that is comprised of multiple data points including, but not limited to, student scores on the annual state exams, graduation rates, student attendance rates, student scores on national college admissions exams and Pennsylvania Value Added Assessment System (PVAAS) growth measures. PVAAS measures the growth each student makes in one school year.

Teacher specific data is based on how each individual teacher's students performed on the state assessments, PVAAS growth of their students, how their students with special needs are progressing on their individual goals and locally developed school rubrics. The final component is elective data, which is determined at the district level but must be approved by the Pennsylvania Department of Education. Elective measures can include district assessments, national assessments, industry certification exams, student projects and portfolios. Elective data



is measured by teachers creating a Student Learning Objective by which they validate and measure the elective data (Pennsylvania Department of Education, 2014).

Professional employees (i.e. tenured teachers) are required to be rated annually and temporary professional employees (i.e. non-tenured teachers) are required to be rated semi-annually. Four ratings are possible: distinguished, proficient, needs improvement or failing. A rating of “distinguished” or “proficient” is considered satisfactory. A rating of “needs improvement” is technically considered satisfactory, but it does result in a teacher being placed on a plan for improvement. If a teacher receives a second “needs improvement” within a 10 year period, it is considered an unsatisfactory rating. Both “needs improvement” ratings must also come from the same employer while the teacher is working in the same certification area. A rating of “failing” is considered unsatisfactory.

### **2.7.1 Effective evaluation of teachers**

As the issue of teacher quality and evaluation has gained prominence, research has been done focusing on what components are important in effective evaluation. This research suggests that there are consistent components that should be present: a quality rating instrument, trained observers, multiple observations of each teacher and evaluation that considers student data (Bill & Melinda Gates Foundation, 2012; Darling-Hammond, 2013). Pennsylvania has designed a comprehensive evaluation system, but developing tools for evaluation does not necessarily ensure effective evaluation. Below are descriptions of the components that should be included in teacher evaluation as it is implemented in schools.

The first component of effective evaluation is the evaluation tool. Teachers need to have a clear understanding of what they are being rated on so the tool needs to set clear expectations

(Bill & Melinda Gates Foundation [Gates Foundation], 2012; Darling-Hammond, 2013). Pennsylvania chose the Framework for Teaching Evaluation Instrument (Danielson, 2013). This tool divides effective teaching into four domains: planning & preparation, classroom environment, instruction and professional responsibilities. The domains are divided into component areas and for each component a description is provided of critical attributes and possible examples. Descriptions are also provided of what performance might look like at each rating level.

The second component is trained observers. To use any evaluation tool correctly, the observers need to complete training and demonstrate competency (Gates Foundation, 2012; Darling-Hammond, 2013). For example, Teachscape is an online software program that provides observers with training on the Danielson model. Observers then complete an assessment where they observe pre-recorded lessons and rate them. Their ratings must meet a certain level of agreement with ratings provided by expert raters. After observers are certified through the program, they must periodically recalibrate to ensure ongoing accuracy. Teachscape is endorsed by the Pennsylvania Department of Education as an observer training tool ([www.pdesas.org](http://www.pdesas.org)).

The third component, use of multiple observations by different observers, also focuses on ensuring accuracy of observations. A single observation may be adequate for the purpose of providing feedback, but when high stakes evaluation is involved multiple observations need to be completed to ensure an accurate picture of how a teacher is performing. The observations should also be conducted by more than one person. For example, if the principal is the primary observer, it is recommended that other individuals also complete observations. Other individuals

might include central office administration, expert teachers and/or instructional coaches (Darling-Hammond, 2013; Gates Foundation, 2012).

The final component is considering observation data in relation to student data. It is better to consider observation data in conjunction with student feedback and student performance on various assessments, including state achievement tests. When considering achievement testing, it is not necessarily only about overall achievement, but taking into consideration students' starting points to determine growth made during the school year (Gates Foundation, 2012; Darling-Hammond, 2013). It is important to note that the research has raised concerns about using student growth for evaluation. As the emphasis on student achievement has increased, some states have moved to using the "value added" model as a tool for teacher evaluation. Value added measures how much a student grows during a school year based on results from testing. Some of the concerns about this model are it does not take into consideration the context, it does not consider factors that are outside the teacher's control, and certain populations can cause the numbers to be skewed (Darling-Hammond, 2013). For example, it does not account for a school where resources may be very limited and many of the children come from a background of poverty. A teacher in this context faces different challenges than a teacher in a more affluent district where resources are ample, and children are sent to school ready to learn. Certain characteristics, such as giftedness, can also skew results. Students who top out on achievement tests have very little room for growth. Therefore, a gifted teacher can appear as not being affective, but the true issue is the test does not provide the room for growth for this population of students. The Pennsylvania model includes value added for teachers in tested subjects but it only constitutes 15% of the rating, and it is a 3 year average.

Another way of gathering evidence of student growth is by using student learning objectives (SLOs). The idea behind SLOs is to gather student evidence on how students are advancing through the curriculum. Examples of how this evidence can be collected include student portfolios, performance-based assessments, state or national tests and locally developed assessments. Teachers set goals and continually monitor progress throughout the year (Danielson, 2013). The Pennsylvania model attributes up to 15% of teacher evaluation to SLOs.

#### **2.7.1.1 Professional development**

The goal of Pennsylvania's Educator Effectiveness model is to reform how school professionals are evaluated as well as professional development (Pennsylvania Department of Education, 2014). This focus on training and development of teachers, based on evaluations, is supported by the literature. Evaluation alone does not necessarily improve teaching practice, but professional development based on feedback from evaluation can improve teacher effectiveness (Darling-Hammond, 2013). "Evaluations should trigger continuous goal-setting for areas teachers want to work on, specific professional development supports and coaching, and opportunities to share expertise, as part of recognizing teachers' strengths and needs" (Darling-Hammond, 2013, p. 99). The following section provides an overview of best practices in teacher professional development.

#### **2.7.1.2 Best practices in professional development**

There are a number of practices that are considered to be effective when delivering professional development that is intended to change instructional practices and impact student achievement. This section provides an overview of the elements that should be included in when planning professional development for educators. The first consideration is duration. Professional

development (PD) should extend over time allowing educators to go in depth with the topic and it needs to include plans for ongoing follow-up (Garet, Porter, Desimone, Birman, & Kwang 2001; Porter, Garet, Desimone, Yoon, Kwang, & Birman 2000). The traditional method of the 1 day conference presented by an expert outside of the classroom has been widely criticized in the literature resulting in an increase in more “reform” style PD activities such as study groups, mentoring and coaching (Garet et al., 2001).

Another component of professional development is educators having some choice in the topic (Porter et al., 2000) so it will be specific to the content the teacher is teaching. This extends to the concept of PD being job embedded (Garet et al., 2001; Hiebert, 1999; Porter et al., 2000), which directly aligns with the theory of andragogy, adult learning, focusing on learning being something that the learners can apply to their everyday situations in the classroom. Job-embedded learning is the practice of educators sharing information, formally or informally, about what works and does not work, trying new practices and reflecting together (Zepeda, 2012). The critical piece here is educators having a opportunities for active collaboration (Garet et al., 2001; Porter et al., 2000). This type of PD provides educators an opportunity to engage in the transformational learning process by discussing and working through common issues.

The *Standards for Professional Learning* (Learning Forward, 2013) provided a guide with the necessary components for effective professional growth opportunities. These standards were specifically titled “professional learning” as opposed to “professional development” to emphasize the active role the educator plays in this process. The standards are divided into seven categories: learning communities, leadership, resources, data, learning designs, implementation and outcomes. The first standard addresses the need for educators to work in learning communities that meet continuously, share collective responsibility and maintain a

focus on their goal(s). The second standard covers the allocation of resources which includes costs of human resources, time for learning, technology, and material resources. This also includes the coordination and fair allocation of resources.

The next standard is data. To know where the needs are in a school, data needs to be collected, analyzed, used in planning, and routinely evaluated. Data sources also need to be rich and varied. The fifth standard covers learning designs. This standard emphasizes the need to incorporate the theories of adult learning when planning for professional learning. The next standard is implementation. Quality professional learning requires sustained effort and this standard talks about the need to plan for on-going support. The final standard looks at outcomes that are based on high expectations. In this standard, professional learning is directly tied to increased student outcomes.

Linda Darling-Hammond (1995) summed it up well when she stated that “Like their students must do, teachers also construct their own understandings by doing: by collaborating, by inquiring into problems, trying and testing ideas, evaluating and reflecting on the outcome of their work” (p. 24). She asserted that by providing educators with opportunities to work on real problems of practice, educators are able to develop a deeper understanding of their practice as well as to develop an increased sense of ownership. The next sections provide specific examples of PD being used in the school communities.

### **2.7.1.3 Models of professional development**

These sections review three specific types of PD being used in the educational setting. These are types of PD that are becoming more common but this is not an exhaustive list of the quality PD options that are now available to educators. The models that will be covered in this section are professional learning communities (PLCs), lesson study and action research.

***Professional learning communities.*** The concept of PLCs is one that embodies many of the elements of effective PD. It is on-going, job-embedded and focused on the specific goals and issues that educators are facing in the classroom. PLCs consist of groups of educators who work together regularly to “identify and apply innovative and effective instructional practices that result in increased student performance” (NAESP, 2008, p. 3). The fundamental idea behind PLCs is that by providing educators an opportunity to connect and collaborate it is possible for schools to improve because of changes in teaching and classroom practices (Harris & Jones, 2010). In the book *On Common Ground: The Power of Professional Learning Communities* (DuFour, Eaker, & Dufour, 2005), the authors summarize PLCs into three big ideas. The first is a shift in focus from teaching to learning. PLCs need to have an intense focus on what they want students to learn, how they will know when students have learned it and what will happen when students experience difficulties. The second big idea is centered on collaboration. A process must be in place for educators to systematically analyze classroom practices. The final idea is focused on results. PLCs determine their effectiveness by analyzing useful data.

A study of PLCs and their impact on school improvement was completed in Wales after the “Welsh Assembly Government introduced a National ‘School Effectiveness Framework’ (SEF) as a way of achieving system-level reform and improved student outcomes for all students” (Harris & Jones, 2010, p. 172). The study focused on the initial pilot schools where PLCs were implemented and the information gleaned from these schools was used to guide implementation efforts throughout the country. By analyzing and comparing schools that successfully implemented PLCs with ones that did not, researchers were able to identify key components for success. Those components were respect and trust among colleagues, supportive

leadership, distributive leadership practices, processes for critical inquiry and continuous improvement, shared vision and a strong focus on learning outcomes (Harris & Jones, 2010).

A number of strengths exist in the PLC model. The first is the intense focus on student learning. In the current era of accountability, initiatives that do not have an emphasis on achievement are not sustainable. For educators, PLCs also provide a learning opportunity that aligns with how adults learn best. There is a clear reason why the learning is important, experiences in the classroom are a major focus and what is being learned by the educators has an immediate and direct impact on the situations they encounter in their classrooms on a daily basis. PLCs also incorporate many of the tenets of effective professional development including that they are job-embedded, on-going and there is frequent collaboration with colleagues. As demands for school improvement increase and funding continues to decrease, schools will have to rethink how they use their resources. The single largest resource in districts is their human resources. PLCs offer a way to capitalize on this resource in a way that can benefit educators by providing an increased level of satisfaction and benefit students by increases in their learning (DuFour et al., 2005).

***Lesson study.*** In the current era of school accountability, it is easy for administrators to find themselves attracted to PD opportunities that promise to quickly and effectively address pressing issues. As was described earlier, the most effective PD happens over time. In other words, the “silver bullet” of PD does not exist or at least has not yet been discovered. In their book, *The Teaching Gap: Best Ideas from World’s Teachers for Improving Education in the Classroom* (2000) authors James Stigler and James Hiebert make the case that teaching is a cultural activity and therefore resistant to change: “The nature of teaching might also help to explain why teaching per se has rarely been the direct focus of efforts to reform education.



Teaching is so constant within our culture that we fail to even imagine how it might be changed, much less believe that it should be changed” (p. 103).

During an extensive study of teaching practices in Germany, Japan and the US, Stigler and Hiebert found that teaching practices in the US have remained relatively stagnant whereas in Japan practices have continually changed and evolved. Part of the reason for this is the belief held by Japanese educators that PD is a continuous process throughout one’s career. One facet of this is the extensive involvement of teachers in lesson study. “By all indications, lesson study is extremely popular and highly valued by Japanese teachers...It is the linchpin of the improvement process” (Stigler & Hiebert, 2000, p. 111). Lesson study is a model of PD that incorporates many of the characteristics of quality PD, but it also has the potential to positively impact the culture of schools.

Lesson study has been a part of Japanese educational culture for many years but it has only been in recent years that it has gained in momentum in the US (Stepanek, Appel, Leong, Mangan, & Mitchell, 2007). The basic premise behind lesson study is it provides “an ongoing method to improve instruction based on careful observation of students and their work”(Lewis, 2002, p. 1). The main elements of lesson study are setting short and long term goals, collaboratively planning a research lesson, teaching the lesson while members of the group observe and gather data, debriefing about the lesson and incorporating suggested revisions for the next teaching of the lesson and finally reflecting and sharing of the results (Lee, 2008; Lewis, 2002; Perry & Lewis, 2009; Stepanek et al., 2007; Zepeda, 2012). There are a number of potential benefits that can occur from lesson study. A few examples are increasing teachers’ capacity to learn collaboratively, increased knowledge of subject matter, and an

increased focus on student learning and understanding based on classroom data (Lewis, 2002; Perry & Lewis, 2009; Stepaneck et al., 2007).

The possible benefits of implementing lesson study in schools are great, but educators pursuing this need to understand the potential pitfalls and lessons learned by districts that have successfully implemented it. Lesson study is not a template or plan that is handed to a faculty to implement within the year. A large case study looking at a 4 year implementation of lesson study in the Bay Area School District (BASD) provides many insights about how to adapt this form of PD into schools in the US (Perry & Lewis, 2009). This study revealed four key conditions that need to be present in order for lesson study to succeed. The first is built in learning opportunities. Lesson study is not a prepackaged plan but rather a process educators go through and continue to learn and refine as they move through it. In the BASD, “from the outset the leaders recognized that their understanding of lesson study was limited; they built *learning opportunities* into their design, hoping that the details of innovation would become clearer with time” (Perry & Lewis, 2009, p. 384), which is what took place throughout the 4 years. The second condition is a professional learning community where teachers are willing to collaborate, the third is distributive leadership and the final one is technical resources. Throughout this process, the BASD teachers incorporated and continually refined a variety of resources such as lesson plan templates, norms for meetings, research materials, outside experts, etc. Unfortunately, one of the biggest challenges in implementing lesson study is lack of time for the meetings and observations that need to take place (Lee, 2008). Other potential challenges include the extra work of lesson study on top of an already heavy work load and for some educators working collaboratively with their colleagues is a skill that needs to be developed before this process has an opportunity to be successful (Lee, 2008). Despite these challenges,

lesson study is an exciting form of PD that can impact student learning and the hurdles can be overcome with commitment and innovative thinking by educators.

**Action research.** One of the purposes of most PD models is to change teaching practices to get better student outcomes. Changing practices is a focus of action research (Corey, 1954; Margerison, 1973; Razfar, 2011), but unlike in the training model (i.e. one day workshops, conferences, etc.) the agent of change is the educator. Action research is a form of PD where “people who actually teach children or supervise teachers or administer school systems attempt to solve their practical problems by using the methods of science” (Corey, 1954, p. 375). In a traditional research model, the researcher attempts to come to a precise conclusion that can be generalized. Action research differs in that the results are more realistic to the situation and have more meaning because they are applicable to a real-life situation (Corey, 1954; Fairbanks & LaGrone, 2006). The process of action research can be broken down into four basic steps that are cyclical in nature (Glanz, 2005; Zepeda, 2012). The steps are select a focus, collect data, analyze and interpret the data, and take action (Glanz, 2005). Throughout these steps the educators are engaging in reflective thought and learning as they participate in discourse which enables them to derive meaning from their experiences and research (Fairbanks & LaGrone, 2006; Hagevik, Aydeniz, & Rowell, 2012).

Challenges are associated with action research that educators need to consider prior to engaging in it. The educator/researcher has to determine how to balance the two roles. For example, collecting data while immersed in the process of teaching can be a daunting task. An obvious solution might be to have another educator come into the classroom as a silent observer and record the data, but the researcher needs to be cognizant that data collection should be done in a manner that is non-intrusive and will not change the dynamics of the classroom (Margerison,

1973). Another consideration is how the educator will remain impartial. “When the teaching role is combined with a research role, there are dual pressures influencing the design and operation of the learning process” (Margerison, 1973, p. 60). Finally, analyzing assessment data can prove challenging (Margerison, 1973, Razfar, 2011). Despite these challenges, action research is a promising model of PD because of the many benefits. It provides educators an opportunity to collaborate on real-life problems, make decisions about practice based on data and develop a mindset of inquiry and reflection (Zepeda, 2012).

## **2.8 POLICY TO PRACTICE**

Thus far, this review of literature contained discussions of what policy is, why certain policy issues rise to prominence, effective teacher evaluation practices and the teacher evaluation policy that was enacted in Pennsylvania. While the policy and resulting teacher evaluation model may be grounded in research, there are challenges that exist when policy moves into practice.

The purpose of policies is to solve problems that have been deemed important, but it is not the policymakers themselves who are charged with implementing the policy. It is the practitioners who are the key problem solvers (Cohen et al., 2007). This relationship can be one of cooperation or conflict. Cohen et al. (2007) identified four factors that tend to determine the outcome of this relationship. Those factors are the nature of the policy, instruments, capability of the practitioners and the environment.

The first factor focuses on two areas. The first is how modest or ambitious the policy is as compared to current practice. If the policy represents only a small departure from current practice, it is more likely to be embraced and implemented well. On the other hand, if it

represents a large change, there are likely to be greater challenges for implementation because it creates greater incapability in those who are implementing (Barbach, 1977; Spillane, 2000). Second is the clarity of the policy (Cohen et al., 2007). Policies are sometimes written in a vague manner due to the need for compromise amongst the policymakers creating the policy, and also ambiguity makes it easier to adapt the policy to fit the local context (Majone & Wildavsky, 1984), but the ambiguity can make it more difficult for practitioners to implement as intended (Cohen et al., 2007).

The second factor takes into consideration the policy instruments. These include incentives, ideas, funding, regulations, etc. Specifically, instruments “offer funds, mandate or forbid actions, create incentives to comply, offer flexibility to adapt to local conditions, and deploy ideas to inform practice” (Cohen et al., 2007, p. 531) This also includes capability building, the next factor, as funding is used to provide resources for this purpose (Cohen et al., 2007; McDonnell & Elmore, 1987). One way to consider this relationship is that instruments bring capability to the implementation, but practitioners also bring with them their own capabilities.

An important aspect of implementation is the capabilities the implementer brings to the process. For example, when considering teacher evaluation an evaluation tool is important, but the understanding of how to use it correctly is critical. Having good resources creates the opportunity, but those resources are only effective if the practitioners know how to use them well (Hanushek, 1996). Capability is partially composed of practitioner values, interests, and skills and knowledge. Depending on a practitioners’ values, a policy may either offend or attract them. If a person is not in agreement with a policy’s aims, the challenge of implementation will be that much greater. Interests refers to whether or not the practitioner has a vested interest in the policy

implementation (Cohen et al., 2007). Skills and knowledge influence the relationships, because a change in practice can require practitioners to do things they simply don't know how to do (Barbach, 1997). This creates a level of incompetence that must be addressed in order for implementation to be successful.

The final factor is environment. An organization can be structured in such a manner that capability to implement a policy is either enhanced or constrained (Cohen et al., 2007). As was already mentioned, policies that cause practitioners to change practice, especially in a major way, create a certain level of incompetence. Organizations can impact how that incompetence is addressed as they enhance capability by enabling and institutionalizing the exchange of knowledge about practice and the interests, values, and knowledge that inform practice”(Cohen et al., 2007, p. 540).

## **2.9 CONCLUSIONS**

Results of the search of the literature suggest that the classroom teacher has the greatest impact on student learning (National Commission on Teaching & America's Future, 1996); yet, ineffective teachers remain in classrooms (Gordon, Kane, & Staiger, 2006). The Pennsylvania Educator Effectiveness Project (EEP) was developed to improve how teachers are evaluated, and provide tools they could use to increase their effectiveness. To date, the efficacy of the EEP has yet to be evaluated in terms of how principals are implementing it and whether those that do implement it have found it to be effective or ineffective in its stated goal to improve the skills of classroom teachers. The present study was developed to gain an in-depth understanding of (a)

factors that led to the passage of the Act 82 teacher evaluation policy, (b) what the policy was intended to do, and (c) how the policy has been translated into actual practice in Pennsylvania.

### **3.0 METHOD**

#### **3.1 RATIONALE FOR STUDY METHOD AND DESIGN**

A mixed method approach was used to explore the research questions with the data collection occurring in two parts. The first part of the collection was an electronic survey and the second was semi-structured interviews. Survey was selected because it allowed for data to be collected from a larger group of participants (Mertens, 2010). This was relevant to the study because of the research that addresses trying to implement reform from the top down in a country that has historically delegated educational decisions to the state and local levels. It is challenging to implement national goals in a country where a national system for education does not exist (McGuinn, 2006). McGuinn (2006) refers to this as the “50/14,000/130,000 problem” in the US. There are 50 state education systems with approximately 14,000 school districts that contain approximately 130,000 schools. Each state has developed a vastly different system of educating students, and many variations exist within the districts. This results in limited capacity for state and federal government to successfully push reform to the local level. Therefore, survey was intended to be an appropriate first step for data gathering, because it was more likely than interviews to reveal inconsistencies between districts.

Upon completion of the survey data, semi-structured interviews of a small subset of participants were conducted. Interviews provide an opportunity to “explore research participant's



perspective about their personal experience with the research topic” (Charmaz, 2014, p. 56). Hochschild (2009) provided several broad purposes for using interviews, one of which was to gain insights into participant attitudes regarding newly developing policy issues. It is relevant to understand personal experience since research suggest that a practitioner’s values, interests, knowledge, and skills can impact how a policy is translated into practice (Cohen et al., 2007). It is for these reasons that the interview method was deemed appropriate for the study.

The structure of the interviews followed a modified version of an in-depth interview format proposed by Seidman (2013). This interview approach was selected because it is focused on perceptions, attitudes, and lived experiences of the participants (Seidman, 2013). The in-depth interview typically consists of three separate interviews, but for this study, each participant was interviewed only one time. The interview followed an in-depth, semi-structured interview with open-ended questions, but all three sections of the protocol were addressed in one sitting. The decision to address all three sections in one sitting was made due to concerns about participation rates.

Seidman (2013) described each of the sections as follows. The first section was designed to focus on the participant’s life history. This provides a context for the participant’s experience based on their previous experiences. For the purposes of this study, it was important to understand each participant’s prior experience with teacher observation and evaluation since it may have influenced their experience and knowledge about teacher evaluation. The second section was designed to focus on the participant’s current experience.

In this study, the focus was on the participant’s current experience working with Act 82 in order to gain an understanding of how the intended purposes identified in the review of literature translated into practice. The final section provided an opportunity for participants to

reflect on the meaning of their experience. As previously mentioned, the method described by Seidman was modified because each participant was interviewed only one time. The decision to modify the format was made in the hopes that it would potentially present less burden for participation. Because principals are busy professionals, asking them to commit to three interviews may have discouraged participation. If additional questions emerged after the initial interview or there was confusion regarding responses, participants were contacted for clarification.

### **3.2 SETTING AND PARTICIPANTS**

The purposeful selection criteria for the surveys was that the principals were serving as a public school (K-12) principal at the time of the study. Their district had to be located in Beaver County or Washington County. Both counties are located in the southwest region of Pennsylvania. Beaver County covers approximately 434 square miles and directly borders the state of Ohio. Washington County borders Beaver County to the south and the state of West Virginia shares Washington's western border. It covers approximately 856 square miles. Both counties are comprised mainly of rural and suburban areas with the closest city being Pittsburgh, Pennsylvania which is located in Allegheny County. Both Beaver and Washington border Allegheny County.

At the time of the study, there were 49 principals serving in a public, K-12 schools in Beaver County and 61 serving in Washington County. The decision to limit the participant pool to these two counties was made for several reasons. The first is the researcher's interest in looking at two sample populations that have a homogenous characteristic within their sample and

then comparing the two sample populations. In this study, the characteristic that districts in each sample have is they are served by the same intermediate unit. Intermediate units (IU) in Pennsylvania were originally established by the Pennsylvania General Assembly to provide cost effective operational and instructional services to school districts in a specific region. They also serve as a liaison between local school districts and the Pennsylvania Department of Education which is why it is being used as a differentiating factor among the participants. This allowed me to look at districts that are served by one IU (i.e. they are in the same county) and also look at how experiences differed if served by a different IUs (i.e. districts in different counties). One of the focus areas was principal preparation for the implementation of Act 82. Since this is a state policy, the IUs have played a role in serving as liaison between the state and school districts, and the IUs are often the providers of training and on-going supports for districts as they go through the process of policy implementation. Considering the difficulties in pushing reform from the top down, I was interested in seeing what kind of variations exist between districts who are served by the same IU, as well as variations that may exist between districts served by a different IUs. Therefore, the decision was made to limit the sample population to two counties served by two different IUs. Beaver County is served by the Beaver County Intermediate Unit which is located in Monaca, Pennsylvania. Washington County is served by Intermediate Unit 1 located in Coal Center, Pennsylvania. Allegheny County was not selected due to my familiarity with the districts and intermediate unit. I have resided and worked in Allegheny County for the past 10 years.

Selected participants had to be in their current position for a minimum of 2 years to increase the likelihood that they have an understanding of the culture of their building and district and have had multiple opportunities to observe the teachers currently employed there.

The next criterion was that participants must have worked with the Act 82 evaluation system for a minimum of one full school year. This criterion was included since some districts had not yet implemented, or were in the process of implementing, Act 82 due to an existing teacher contract. For the purpose of this study, it was important that the participants have sufficient experience with the evaluation system to provide an in-depth and stable description of their experiences.

The original goal for the study was to have a pool of diverse respondents participate in the interview. From the pool of participants who indicated a willingness to participate in a follow-up interview, the hope was to create a group representing as many diverse environments as possible. Once the groups were determined, I planned to use proportional quota sampling to select an equal representation of participants from each group. Participants were to initially be divided into groups based on level they serve (elementary, middle and high school) and school district. It was hoped that the initial grouping would provide the opportunity to look for differences among grade levels and the various school districts. Additionally, demographic information about the districts was to be taken into consideration. Districts were to be grouped with other similar districts based on total size of student population, percentage of economically disadvantaged students and academic achievement. The percentage of economically disadvantaged students would be based on the student poverty concentration as reported by the Pennsylvania Budget and Policy Center. Academic achievement was based on Pennsylvania System of School Assessment (PSSA) data for K-8 buildings and Keystone data for high schools as reported by the Pennsylvania Department of Education (PDE).

PSSAs and Keystones are state mandated achievement tests that are administered on an annual basis in the spring. Once these groups were developed, representatives from each group would be selected and an attempt would be made to have an equal representation of male and

female principals. Unfortunately, the response rate to the initial survey was very low with only 10 people responding and 1 indicating she would be willing to interview. Additional efforts, detailed in a following section, were done to try to increase participation rate. By the end of the study, I was able to obtain only five additional interview participants so all respondents who were willing to interview were included.

When completing qualitative research, the literature indicates that no definitive guidelines exist for determining sample size when working with a non-probabilistic sample group. This means the researcher should consider relying on “saturation” to determine the final sample size. Saturation refers to the point in which no new information or themes are observed in the data (Guest, Bunce, & Johnson, 2006). In a review of research that does provide some guidelines for determine sample size, Guest et al. (2006) found recommended samples sizes ranging from five to several hundred. Therefore, for this study, the goal was to have a minimum of five follow-up interviews conducted, with a maximum number not be determined until the interviews were underway and saturation was reached. Due to response rate issues, all respondents were included and saturation was not reached.

### **3.3 ETHICAL CONSIDERATIONS**

The main purpose of this research was to learn how principals understand the intended purposes of Act 82, how they were prepared for implementation and how the intended purposes have been translated through the implementation process. A major part of this research was to gain an understanding of principals’ attitudes, perceptions, and lived experiences with Act 82. To gain authentic feedback, it is critical that participants feel comfortable providing open and honest

responses. Procedural safeguards were put into place to assure the confidentiality of the interviews and anonymity of the participants. The introduction letter sent prior to the survey and interview explaining the purpose of the research contained a statement that participation was voluntary, described any risks associated with participation, described the survey and interview formats, and defined method to assure confidentiality.

Participants were also made aware that if they choose to participate, they would have the option to refuse to answer any question and they could withdraw from participation at any time (Seidman, 1991). Prior to the start of the interview, this information was read aloud to them, and they received an emailed copy of the consent form since all participants chose phone interview. Participants were then asked to affirm their consent to participate. Permission to audio record the interview was also requested, and all participants agreed to be recorded. All audio recordings were maintained in a secure location. All identifying information regarding the participant or school was removed from the transcribed results. In the transcription, participants were labeled by a number. A separate list was created matching each number with participant name and school. I maintained the list on a password protected computer and only referenced it to assist in understanding the findings. No identifying information was reported.

In the article, *Pearls, Pith and Provocation: Issues of Representation Within Qualitative Inquiry*, Mantzoukas (2004) asserted that when completing qualitative research the researcher needs to identify the “paradigmatic positioning” of the interviewer including identifying ontological assumptions to address accuracy and representation. This is assuming the researcher is adhering to a non-positivist paradigm. Therefore, I explicitly defined my own assumptions as the researcher to provide the reader a lens in which to understand and interpret the findings.

My belief system falls within the constructivist paradigm. The ontological assumption is that reality is socially constructed as opposed to an absolute reality (Mertens, 2010). My view of the data collected from the participants was that their responses were based on their understandings and knowledge as influenced by the social milieu of their professional lives. Their understandings would also be fluid as they continued to have experiences with the teacher evaluation process. For this reason, I excluded administrators with less than 1 year of experience with the target evaluation process. Participant beliefs and understandings may be impacted by the interview process as they reflect on their experiences and go through the process of describing them. Finally, my own experiences and thoughts about the new evaluation model are described in the reflections in Chapter 5. I remained aware of my own assumptions as I worked through the coding and analysis of the data. This process was facilitated by tracking my own reactions to the data as recommended by Sunstein and Chiseri-Strater (2007). The authors recommended that, as the researcher is coding, notes should be taken to indicate where the researcher felt surprised, intrigued, and/or disturbed by the data. This provides insight into the researcher's assumptions, value system, and beliefs.

### **3.4 DATA COLLECTION PROCEDURE**

Data for this study was collected through an electronic survey and a semi-structured interview with open ended questions.

### **3.4.1 Study protocol**

The study proceeded in the following manner. Permission to use human subjects was obtained from the University of Pittsburgh Human Research Protection Office pursuant to the U.S. Federal Government Department of Health and Human Services (2009) regulation 45 CFR § 46.10, which states the probability and magnitude of harm or discomfort anticipated in the research should not be greater in and of themselves than any ordinarily encountered in daily life, or during the performance of routine physical or psychological examinations or tests. An initial invitation to participate in the survey was sent via email to all public school principals in Beaver and Washington counties (Appendix B). Included in the survey were questions asking if the participant met the eligibility criteria of being in their current position for a minimum of 2 years and having worked with the evaluation tool for a minimum of 1 year. If they did not meet these criteria, they received a message thanking them for their willingness to participate, and they were exited from the survey. For participants who were eligible, they were asked at the end of the survey if they would be willing to participate in a follow-up interview.

Reminder emails were sent to principals who had not responded. Qualtrics web-based service was used for design and electronic delivery of the survey. The University of Pittsburgh Human Research Protection requires, in most cases, that researchers utilize Qualtrics for web-based survey research and therefore provides the service free of charge.

The last question in the electronic survey asked if respondents would be willing to participate in an interview. If they were willing, they were asked to provide their contact information. Participants who chose to complete the interview were given the option of being interviewed over the phone or in person. All participants chose to be interviewed via the phone. Prior to the interview, I contacted the participants either by phone or email to schedule a date and



time for us to talk. Before the interview they received, via email, the Invitation to Participate and Consent Form (Appendix C). They were asked to indicate their consent to participate. Since they received the consent form via email, they responded to the email indicating their consent to participate. Prior to the start of each interview session, participants were asked for permission to have the interviews audio-recorded.

All participants granted permission and interviews were recorded in their entirety. When I started the audio recordings, I asked participants a second time if they gave permission to be audio recorded so the granting of permission would be recorded. They were told ahead of time that I would be asking a second time once the recording was started. Two means were used to audio record the interviews. The first was a basic Sony voice recorder. The second was an iPhone using the app RevRecorder. This app functions as a voice recorder, but it also provides an option to send the recording in for transcription. All of the interviews were submitted to be transcribed for a fee. The electronic transcription was emailed to me. I then review the transcripts for accuracy. The rationale for using audio-recording, as opposed to note taking, eliminate the possibility that I might unconsciously make slight modifications to meaning. Had a participant denied permission to audio record, I would have taken hand written notes.

The initial round of surveys yielded a very low response rate with 10 survey responses and 1 respondent indicating willingness to do an interview. In an effort to increase response rate, I requested permission from the University of Pittsburgh Human Research Protection Office to reach out to potential participants individually. Permission was granted to call and email principals using the Follow Up Script for Low Response Rate (Appendix E). I was also granted permission to speak to principals at a principal meeting at the intermediate units. Therefore, I either emailed or called potential participants using the script, and I attended a principal meeting

at the Beaver Valley Intermediate Unit. This effort yielded a greater response rate details of which are included in Chapter 4.

### **3.4.2 Pilot of the instruments**

Yin (2011) stated that pilot studies assume the role of a laboratory that enables the observation of the phenomenon under study from various angles and approaches on a trial basis. Pilot testing of the survey and interview protocol (Appendix D) is an activity designed to refine the data collection process relative to both the content of the questions and the procedure to be followed (Yin, 2011). A pilot study serves as a “full-scale dress rehearsal” (Hulley, 2007, p. 277) employing the study’s survey and interview protocols and participant observations. The goal of the pilot test in this study was to determine whether the survey and interview questions and data analysis would be sufficient in achieving the overall objectives of the research questions. A pilot study has two major objectives. The first objective is to administer the questionnaire to a small sample of people to refine the questions, as well as observe the procedure. The second objective is to initiate preliminary data analysis to determine the effectiveness of the resolution of data.

For the purposes of this study, both instruments were piloted with participants from the Doctoral Study Group in Administrative and Policy Studies-School Leadership led by Dr. Cynthia Tananis. This group serves to provide doctoral students the needed support and tools to complete the comprehensive exam and dissertation process. I have been a member of this group for the past several years. Members of this group were asked to take the survey if they also met the requirements of currently serving in the role of principal in a public school and had been in this same position for a minimum of 2 years. Once they had completed the survey, they were asked to provide feedback on clarity of the questions, suggest additional questions and provide

general feedback. The interview instrument was sent out to all members of the study group for feedback on clarity of questions. Members of the group provided both written feedback and verbal feedback during a study group session. The instruments were revised based on the feedback received. Participants in the pilot study process were not included in the main research project, nor was the data they generated included in the data analysis.

Creswell (2012) acknowledged that pilot testing may be required because no amount of intellectual analysis can substitute for testing the validity of interview questions designed to generate information responsive to the research questions. Piloting the instruments for this study helped ensure the clarity, user-friendliness, wording, and meaning of the survey and interview questions.

### **3.5 RESPONSE RATE**

#### **3.5.1 Survey distribution and response rate**

The survey was initially sent through the Qualtrics system. Potential participants received an anonymous survey link via an email. Currently, there are 49 principals serving in a public, K-12 school in Beaver County and 61 serving in Washington County. The survey was sent to 45 principals serving in Beaver County and 58 principals serving in Washington County. There were several reasons for the discrepancy in numbers. Multiple attempts were made to find contact information for each principal including searching district websites, general web searches and contacting the schools directly. Despite these efforts, I was unable to include several principals in the study. In a couple of cases, it was evident that some principals had moved out

of their positions based on information posted on the website. If there was a new principal in the position, (s)he would not have met the eligibility criteria for the study. I was unable to locate emails for several principals despite multiple attempts. In one instance, I was personally aware that one principal had moved to another district outside of the counties being studied.

In the survey, the first few questions probed each principal's experience to determine if they met the criteria for the study. The criteria set forth was that selected principals would have been in their current position for a minimum of 2 years. This was included to increase the likelihood that they had an understanding of the culture of their building and district and had multiple opportunities to observe the teachers currently employed there. Therefore, principals who did not meet this criteria were excluded from the survey if they did not respond in the affirmative.

The first distribution of the survey was sent in early August 2016, with two reminders sent to unfinished respondents. There were 10 participants who responded during this phase. Six were from Washington County and four were from Beaver County. Nine respondents completed the survey and one was unfinished. Four emails bounced back. It is possible that others were not delivered as well. I discovered later in the data collection phase that some districts have an email program that requests a response from the sender when the email is from an unrecognized account. It is possible that while some of those emails did not bounce back, they did not reach the person.

The initial response was 8%, which may have been because the survey was sent close to the start of the school year, which is an especially busy time for principals. Therefore, the decision was made to wait and do a second distribution. The survey was sent out again at the beginning of October and yielded one more response. At this point, the decision was made to

reach out to participants personally in an attempt to gain more respondents. I applied to the University of Pittsburgh Human Research Protection Office for permission to call or email principals directly to request their participation. I also requested permission to attend a meeting at the intermediate unit to introduce my study to principals. Permission for both of these activities was granted.

During the month of October 2016, I emailed or called potential respondents using the script located in Appendix E. I also attended a principal meeting at the Beaver Valley Intermediate Unit where I introduced my survey and requested their participation. The survey was distributed again at the end of October with two reminder emails sent in early November. After the final distribution, 27 surveys were completed equaling a response rate of 26%. During a conversation with an interview participant, she indicated that they have been inundated with surveys recently. This may have been a factor in the initial low response rate as well as time of the school year. Of the surveys completed, 59% were completed by respondents in Washington County and 41% were completed by Beaver County respondents.

### **3.5.2 Interview response rate**

In the survey that was sent out, the last question asked respondents if they would be willing to participate in an interview. If they were willing, they were asked to provide their name and contact information. During the August 2016 distribution, one principal responded that she would participate and provided contact information. A phone interview was completed in August. In response to the October 2016 distribution, three additional principals indicated that they were willing to be interviewed. They were interviewed in November. Two additional principals responded to the email that was sent, and they were also interviewed in November.

All principals were interviewed over the phone. I had initially hoped to interview in person, but found that principals preferred the phone interview, which may have been due to their restrictive schedules. The consent form was emailed to them, and they all responded indicating their agreement to participate. All interview participants agreed to be audio recorded. I used two devices to record. One was an iPhone using the app Rev Recorder. This app allowed me to send the recording in to be transcribed. A Sony digital voice recorder was used as a backup. Several of the transcripts were returned with areas that were inaudible. Therefore, I reviewed the recordings and transcripts to ensure accuracy.

### **3.6 DATA ANALYSIS**

#### **3.6.1 Data management**

Data collected from the electronic survey was managed and stored in Qualtrics. Some of the information collected from the survey was entered into Intellectus Statistics software for the purpose of analysis. Data collected from interviews was transcribed to assure an accurate representation of the participant's thoughts (Seidman, 2013). The audio-recorded interviews were recorded using Rev Voice Recorder, which is an online app that audio-records and provides a written transcript. No personally identifiable information was included in the audio-recordings. Participants were identified via a randomly assigned number. The conversations were transcribed in their entirety because to do otherwise could lead to inaccurate decisions about what information was important and what was not. At the conclusion of all interviews, participants were asked if they might be contacted at a future time for clarification of an answer

or if further information is needed. Multiple methods were used to analyze the data collected during this study. During the coding and analysis phases, key phrases were copied and pasted into documents with each participant's identifying number to provide additional anonymity for the participants.

### **3.6.2 Analysis method**

The research questions were focused on the perceptions, attitudes, and lived experiences of school principals who had been involved with the implementation of Act 82. Survey and interview questions (Appendix C & D) guided an exploration of how the policy has translated into practice. The conceptual framework was focused on the policy to practice research which is divided into four sections: ambiguity versus clarity, instruments, capability and environment. Data analysis occurred by analysis of survey data using descriptive and inferential statistics as detailed later in this section. Survey and interview data were completed through the use of coding using this framework as a guide.

RQ1 was an effort to uncover what principals understand about the intended purposes of Act 82. This falls under the area of ambiguity versus clarity. In vivo coding was used to analyze participant answers. This type of coding was chosen because it captures the participant's voice by focusing on choice of language. Data came from the survey and transcripts of participant interview answers to questions specifically about intended purposes. Coding occurred by reading through participant answers and highlighting words and phrases that highlighted participant understanding of intended purposes. The highlighted words/phrases from all participants were then condensed into one document to facilitate analysis. The data were analyzed for emergent themes by looking for commonalities in what participants identified as

intended purposes. Themes that emerged were noted. The data was also analyzed for variations in participant answers. Once this was completed, the original transcripts were reviewed a second time for any additional data that appeared to support emergent themes or variations. Once themes and variations were identified, they were compared to actual intended purposes. This analysis was used to draw conclusions about the level of clarity/ambiguity of the reasons for the policy.

RQ2 was an exploration of principals' perceptions related to how they were prepared for implementation of the policy. This falls under the area of instruments (i.e. training) and capability (i.e. familiarity or prior experience with the 4 domains). Data collected in this area came from survey and interview questions specifically focused on instruments and capability and were coded into those two main categories. In the area of instruments, survey answers were analyzed to determine if a relationship existed between the type and duration of training provided and participants' feelings of preparedness. In the area of capacity, I looked at participants' familiarity with the evaluation tool to see if a relationship existed between prior experience with a similar tool and feeling prepared for implementation.

Initially, survey answers were to be analyzed using cross tabulation with chi square to see if a relationship existed between the type and duration of training provided and their feelings of preparedness. The planned Chi-square analyses were conducted, however, due to insufficient cell sizes, the analysis could not be used. For this reason, the inferential analyses focused on intermediate unit training effectiveness and duration, along with perceptions of preparedness, prior experience, and use. Additionally, differences between the two counties on each of these variables was assessed.



A Pearson correlation analysis was conducted to determine if there was a relationship between participants' feelings of preparedness and Intermediate Unit training effectiveness, prior use of the tool, and prior experience with the Danielson model. A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in Intermediate Unit training effectiveness between the reported levels of preparedness. A Chi-Square Test of Independence was conducted to examine whether county and preparedness were independent. A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in Intermediate Unit training effectiveness between the counties. A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in Intermediate Unit training duration between the counties.

Coding of the interview data began by reviewing the transcripts with the first round focusing on references to instruments and capacity. Specifically, in the area of instruments I was looking for references to trainings and in the area of capability I was looking for participant's prior experience with the evaluation tool. The interview transcripts were reviewed and any reference to a policy instrument or capability was highlighted and copied into a separate document. Each reference from the categories listed above was coded as instrument, capability or a combination of both. A second round of deductive coding occurred by reviewing the comments made about instruments and capacity. Analysis was done comparing participants' experiences in both areas with their feelings of being prepared. Specifically, I was looking for evidence that increases in capability, effective instruments and/or a combination of both would lead to higher levels of feeling prepared. This information was also analyzed to see if it suggested differences between support/training provided by the two I.U.s that serve these districts. This process was repeated when looking at capability, except analysis also included

looking for variations and consistencies between individual participants. Throughout this process, notes were written to capture overall impressions, any data that was surprising, and reflections on analysis.

RQ3 was intended to provide focus on how Act 82 has been implemented in schools. This falls into the category of environment. A priori coding was used and was divided into two main categories: effective supervision practices and professional development. In the supervision practices, I looked for examples of effective practices i.e. effective tool, trained observers, and multiple observers for each teacher. Data provided by the participants was categorized into one of these three areas. In the category of professional development, I looked for evidence of embedded, on-going professional development, and evidence that data from teacher evaluations was taken into consideration when planning for PD. Data were categorized into these three areas: PD is embedded into the job, PD is on-going, and data from teacher evaluations considered when planning for PD. I also looked for any unanticipated responses that fell outside of these categories. Once coding was complete, the data were analyzed to look for evidence that effective supervision practices and professional development were or were not occurring and variations that existed between participants and districts.

Survey data was collected and managed using Qualtrics. Some of the survey data was input into the Intellectus Statistics software for the purpose of analysis. Interview data were collected via audio recording and transcribed into a word document. The data were managed using word processing; the first result was a compilation of all data received (Table 1). For RQ1, key phrases from each participant were copied and pasted into one document. For RQ2 and 3, separate documents for each category were developed. Key phrases were copied and pasted into

those documents with each participant's identifying number. In all areas, I also looked for any unanticipated responses, analyzing them to determine their meaning.

**Table 1.** *Data Analysis Chart*

Research Questions	Data Source & Collection	Data Collection Items	Alignment with Literature	Data Analysis
RQ1: What are school principals' understandings of the intended purposes of Act 82?	Survey and Semi-structured interviews of currently practicing K-12 public school principals in Beaver County	<i>Survey Question:</i> 11 <i>Interview Questions:</i> 13	Multiple Streams Theory  Policy to Practice	Data was coded by determining emergent themes and analyzing their alignment with actual intended purposes.
RQ2: How have principals been prepared for the implementation of the Act 82 Teacher Effectiveness policy?	Survey and Semi-structured interviews of currently practicing K-12 public school principals in Beaver County	<i>Survey Question:</i> 6, 7, 8, 9, 10, 11 <i>Interview Questions:</i>	Policy to Practice	Data was coded into two areas impacting how policy is translated into practice: instruments and capabilities (practitioner prior experience, familiarity, skills & knowledge).
RQ3: How have principals implemented the Act 82 Teacher Effectiveness policy in their buildings?	Survey and Semi-structured interviews of currently practicing K-12 public school principals in Beaver County	<i>Survey Question:</i> 5, 6, 11, 12, 13, 15 <i>Interview Questions:</i> 6, 7, 8, 9, 10, 11, 15, 16	Policy to Practice  Effective Teacher Evaluation  Best Practice in Professional Development	1.) Data was coded to identify best practices in effective evaluation: effective tool, trained observers, multiple observations & evaluation tied to student data. 2.) Data was coded by professional development practices in use as related to evaluation data. The first level of coding looked at if PD is based on evaluation. The second round coded for PD based on best practices: duration, choice in the top, job embedded and collaboration.

### **3.7 ACCURACY AND REPRESENTATION**

This section is divided into two subsections relating to the issues of accuracy and representation. The first part focuses on these issues as they relate to the participants. The second part reflects on role of the researcher.

#### **3.7.1 Participant issues**

This study was based on survey and semi-structured interviews of participants. All of the data were self-reported by the participants. Accuracy of the data is contingent on the honesty of the participants. This does not imply that participants might be purposively deceptive, but rather they may be unsure about an answer or not realize that they do not know the answer (Mertens, 2010). Other factors may also impact participant's ability to answer accurately such as memory loss, confusion, fatigue, or other outside factors (Charmaz, 2014). Three elements were included in the interview protocol to increase the accuracy of the data.

##### **3.7.1.1 Interview sections**

The interviews were structured into three parts. By dividing the interview into three sections, understanding of the participant's background experience (i.e. context) with teacher evaluation was enhanced. As well, asking questions about the meaning each principal attached to the current experiences was explored. This enabled a check for internal consistency in answers. If inconsistencies arose during the interview, a probe was initiated to try to determine what was causing the inconsistency.

### **3.7.1.2 Interview scheduling**

The second element was the scheduling of the interviews. As the researcher, the schedule was carefully constructed to occur at times when the participants were more likely to feel rested and able to focus. School principals have complex jobs requiring many hours. The goal was to schedule at a time outside of the work day when other distractions would be less likely.

## **3.8 SUMMARY**

Chapter 3 was a discussion of the construct of the present mixed method study. The discussion included a review of the research questions, the intent of them, and how they will be implemented. The procedure for the collection of data was summarized as well as the procedure the analysis of both quantitative and qualitative data resulting from the collection of data. Chapter 4 contains a summary of the results based on the data collected. Chapter 5 contains the analysis of the study.

## **4.0 DATA COLLECTION RESULTS**

### **4.1 INTRODUCTION**

The literature suggests that one of the most significant factors impacting a child's education is a quality teacher. The Pennsylvania Educator Effectiveness Project (EEP) was developed to reform how teachers are evaluated and provide the tools with which they can increase their effectiveness. This study was divided into two parts. Part 1 was an exploration into factors that led to the passage of the current evaluation system and the intended purposes. This analysis was completed in the Chapter 2 Review of Literature using John Kingdon's Multiple Streams theory (Kingdon, 2011). Part 2 focused on principals' experiences with the policy during the early years of its implementation. Data for Part 2 was collected via electronic survey and semi-structured interviews

### **4.2 PARTICIPANT INFORMATION**

#### **4.2.1 Survey participants**

There were 28 principals who responded to the electronic survey, but only 27 completed it. Below are tables and descriptions of the participants based on their survey responses. Principals

were asked what grade levels their buildings serve. The question indicated that they should select the answer that “best” described their building since there are a variety of configurations that exist. There was a fairly even split between elementary principals (44%) and principals serving at the secondary level (56%) as shown in Table 2. For purposes of this study, secondary refers to buildings serving any combination of grades 5-12.

**Table 2.** *Respondent Demographics*

Grade Level of Building	Number of Respondents	Percentage
Elementary (K-6)	12	44%
Middle School (5-9)	6	22%
High School (9-12)	5	19%
Middle/High School (5-12)	4	15%

The next question asked principals how long they had served in their current position. If they answered less than 1 year, they were exited from the survey. The majority of the principals (79%) reported having been in their current position from more than 3 years as shown in Table 3.

**Table 3.** *Number of Years in Current Position*

Years In Current Position	Number of Respondents	Percentage
2 Years	6	21%
More than 3 years	22	79%

The majority of respondents were from Washington County (59%), but the group had a fairly even split as shown below. One respondent did not answer the question as shown in Table 4.

**Table 4.** *Location of School District by County*

County	Number of Respondents	Percentage
Beaver County	11	41%
Washington County	16	59%

#### **4.2.2 Interview participants**

This section contains a description of the principals who participated in the interview section. The descriptions include participant background, gender, building level where they currently serve as principal, size of the district, student economic information, and academic achievement. District and building level data for economically disadvantaged rates and academic achievement was found in the School Performance Profile (SPP) reports for the 2015-2016 school year (Pennsylvania School Performance Profile, 2016). The percentage of students identified as economically disadvantaged was based on the number of students who qualified for free or reduced lunch. Federal qualifications for free or reduced lunch are as follows: a family of four with an annual income of \$23,850 qualified for free lunch; an annual income of \$44,123 qualified for reduced lunch (United States Department of Agriculture, 2016). The SPP provides building level information about the student population and an academic score based on a



multitude of factors. The SPP is heavily weighted by academic achievement, specifically PSSA results and Keystone exams. The SPP is a component of teacher evaluation.

#### **4.2.2.1 Interview participant background**

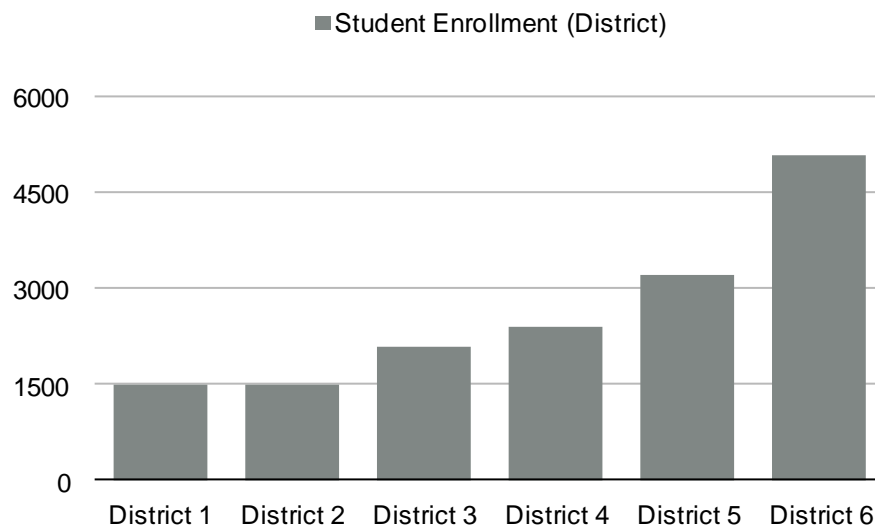
The interview participant group was comprised of four women and two men. There was one high school, one middle school, one middle/high school, and three elementary principals. Four serve in schools located in Washington County and two are in Beaver County. The participants had varying years of experience in education and how long they had served in an administrative position. All, except two, had over 15 years of experience in education. One participant had 9 years and another had 11 years of educational experience. In regard to number of years in administration, one participant had 2.5 years, one had 4 years, two participants had 6 years, one had 17 years, and the other had 20 years.

#### **4.2.2.2 Participant building and district information**

This section provides information about the school buildings and districts the interview participants serve in. As stated earlier, the goal was to have participants selected from purposive sampling. All of the interview participants were from different school districts. The districts and buildings where they work are a variety of sizes. The smallest building serves around 260 students whereas the largest building serves over 1,000 students. The buildings in the middle serve approximately 430 to 650 students.

The districts also vary in size. The smallest two districts serve approximately 1,500 students. Two of the districts serve between 2,000-2,500 students. One district has a population of about 3,200 and the largest serves around 5,000 students. While variation exists between the participants' buildings and districts, there was not a great variation considering that all of the

participants, with the exception of one, have an ED percentage rate between 25-42%. The one exception has a building ED population of 20% qualifying.



**Figure 1.** *Student Enrollment*

In the area of student achievement, the participants' buildings were more similar to each other than dissimilar. This conclusion is based on their SPP ratings. The PSSAs are given to students in grades 3-8 in the two subject areas of mathematics and English Language Arts (ELA). In addition, students take a Science PSSA in fourth and eighth grades. The Keystone exams are considered "end of course" exams given when a student completes Algebra, Biology and Literature.

The SPP takes into consideration how many total students were proficient on these exams and how much growth a district has demonstrated. Forty percent of the SPP score is based on number of students who demonstrated proficiency on the PSSAs and/or Keystones, the percentage of students earning an industry certification, and SAT/ACT benchmarks. The third grade English Language Arts PSSA score is more heavily weighted than the other PSSA scores

because of the importance of being able to read by the time a student completes this grade. Another 40% of the score is based on indicators of academic growth. This refers to the building's ability to impact change from year- to-year for each cohort of students as measured by the PSSAs and /or Keystones. Attached to each SPP score is a color: red, yellow, green, light blue or dark blue. Below is a table of the colors and corresponding scores (Table 5).

**Table 5. Building Scores**

Color	Score
Dark Blue	90-100
Light Blue	80-89.9
Green	70-79.9
Yellow	60-69.9
Red	less than 60

All of the buildings represented by the participants had building scores in either green or light blue. This indicates that the buildings are achieving a proficient level or above. It also indicates that their students are making a year's worth or more of academic growth in one school year. Overall, the districts are fairly similar in academic achievement.

The following section is a summary of the data collected to answer the research questions and an analysis of the data. The data is presented by research question.

### 4.3 DATA COLLECTION AND ANALYSIS OF RQ1

*RQ1: What are school principals' understandings of the intended purposes of Act 82?*

RQ1 was an effort to uncover what principals understand about the intended purposes of Act 82 as compared to the actual intended purposes of the legislation. This falls under the area of ambiguity versus clarity. The stated goal of Educator Effectiveness is “To develop educator effectiveness models that will reform the way we evaluate school professionals as well as the critical components of training and professional growth” ([www.education.pa.gov](http://www.education.pa.gov)). In Chapter 2, a historical review of events leading up to the implementation of Act 82 was explored in an effort to understand fully the intended purposes beyond the stated goal. As described in Chapter 2, Act 82 was implemented as a result of funding from the Race to the Top (RTTT) competitive grant program.

RTTT was traced back to a report from the Brookings Institute. The Brookings Institute published a report in 2006 titled *Identifying Effective Teachers Using Performance on the Job* authored by Robert Gordon, Thomas Kane, and Douglas Staiger. The report focused on increasing teacher quality through five recommended actions. Through an analysis of historical documents, with an emphasis on the Brookings Institute report, Race to the Top documents, and Act 82, I identified the following purposes of Act 82: ensure high quality teachers in classrooms, ability to remove ineffective teachers, evaluation systems with multiple measures, use of student achievement data as part of the evaluation process, and focus on teacher professional development and training. More detail is provided in Chapter 5 about how the intended purposes were identified. The following analysis compares principal responses about intended purposes to the ones identified above.

In Vivo coding (Saldana, 2013) was used to analyze participant answers. This type of coding was chosen because it captures the participant’s voice by focusing on choice of language.

Data to answer this question came from the survey and transcripts of principal answers to a specific question about intended purposes. The specific question is listed below.

Survey and Interview Question: *What do you think policymakers were hoping to accomplish by creating the teacher evaluation system that is now in place?*

Coding occurred by reading through participant answers to the above question and highlighting words and phrases in an effort to understand participant interpretations of intended purposes. Saldana (2013) recommended that during this process the researcher attune oneself to “words and phrases that seem to call for bolding, underlining, italicizing, highlighting or vocal emphasis if spoken aloud” (p. 92).

The highlighted words/phrases from all participants were condensed into one document to facilitate analysis. The data was analyzed for emergent themes by looking for commonalities in what participants identified as intended purposes. Themes that emerged were noted, and the data was analyzed for variations in participant answers. The transcripts were then reviewed a second time for any additional data that appeared to support emergent themes or variations. Once themes and variations were identified, they were compared to actual intended purposes. This analysis was used to draw conclusions about the level of clarity/ambiguity of the reasons for the policy. The data suggested that the participant's understanding of the intended purposes generally aligned with the intended purposes.

Three themes are discussed in the following sections: accountability, effective evaluation tool, and consistency. The predominant theme that emerged was accountability. There were two sub-themes within accountability: focus on student achievement, and identifying ineffective teachers. Another theme that emerged was development of an effective evaluation tool. This appeared to align with the intended purpose of ensuring high quality teachers, but it needs to be

noted that there was a divergence in the area of teacher growth. Participants alluded to the evaluation tool itself helping promote teacher growth as opposed to professional development being the factor in that growth. In this case, they were referencing the 50% observation section that is based on the Danielson framework. In Act 82, the goal specifically references teacher growth in regard to professional development and training. In the responses, only one principal referenced “continuing education.” Finally, creating consistency between districts and buildings came up in several responses and two participants indicated they were not sure of the purposes.

Table 6 shows the intended purposes and principal perceptions of intended purposes. The predominant theme is accountability. Under “Intended Purposes of Act 82,” the specifics of what the policy was intended to do are listed, but in parentheses it is noted which purposes are accountability measures. Under principal perceptions, accountability is listed as its own purpose since that specific wording choice was either used by a majority of the participants or referenced in a response. This suggests alignment with four of the intended purposes even though participants may not have specifically described the items in those four corresponding areas.

**Table 6.** *Comparison of Intended Purposes and Principal Perceptions*

<b>Intended Purposes of Act 82</b>	<b>Principal Perceptions of Intended Purposes</b>	<b>Percentage of Participants Identifying Purpose</b>
Ensure high quality teachers (Accountability)	Effective tool to help develop quality teachers (Teacher Growth)	47%
Ability to remove ineffective teachers (Accountability)	Identifying ineffective teachers (Accountability)	16%
Evaluation system using multiple measures (Accountability)	(*Identified in only 1 response)	—
Use of student achievement in evaluation (Accountability)	Increase student achievement (Accountability)	16%
Focus on professional development and training	(*Identified in only 1 response)	—
	Accountability	68%
	Consistency	16%
	Unsure of purpose	11%

In the following sections, each theme identified in the data is described in more detail.

#### **4.3.1 Accountability**

The theme of accountability was the most prevalent throughout the responses. One survey question and one interview question asked participants specifically about what they felt were the intended purposes of this policy. The survey question was designed as an open-ended response.

Nineteen principals provided a response in the survey. Of the 19 responses, 8 answers included either the word “accountable” or “accountability.” Four of the six principals who were interviewed also specifically used some form of the word accountability. Overall, 42% of the principals responding to this question included a form of the word accountability in their responses. Interview responses are not included in this overall calculation. It is likely they answered the survey question in a similar fashion so to include them would potentially cause them to be counted twice.

The questions where the word “accountability” was not used explicitly were analyzed to determine if the principal’s answer inferred some type of accountability. In the survey response, there were an additional five principals (26%) who did not use the word accountability specifically, but referenced it. This increased the total percentage of principals referencing accountability to 68%. Two answers referenced that the purpose was to identify unsatisfactory teachers. Two answers referenced a focus on student achievement. One additional answer stated that the intended purposes was to “get the most out of teachers.” Including these additional answers brings the percentage of principals who referenced accountability to 68% of the principals responding to the question. Again, the number only takes into account the survey responses.

In the interview responses (Table 7), two of the principals did not use the word “accountability” specifically. One principal did reference being held accountable in her response. She provided this description:

Coming from out of state, and hearing stories from my nieces when they were in school, I'm sure the legislators thought these teachers don't do anything we need to put stuff into



place to make sure they're doing their jobs. When I was in another state my nieces would tell me, early May, that they've already turned their textbooks in for the year.

The other principal also referenced accountability in her comments, “I think policy makers wanted to ensure that teachers were reflecting and working on their practice.”

**Table 7.** *Responses Referring to Accountability*

<b>Responses Referencing “Accountability”</b>
“more effective system to hold teachers accountable for the instruction”
“I believe there was also an accountability factor by policymakers to improve and raise the bar for school improvement purposes. “
“more accountability for teachers”
“To increase teacher accountability”
“an accountability measure for effective teaching and learning.”
“They may have been trying to hold professionals more accountable. “
“The policy makers were looking for greater “accountability.”
“My sense is that there is an accountability piece to it trying to make people more accountable.”
“motive to try to build some more consistent accountability across the state”
“trying to make people more accountable”

Within the theme of accountability, there were two sub-themes that emerged. The first was the use of student achievement within the teacher evaluation system. Only four participants

referenced this specifically and one alluded to it. Three of those responses were from the survey equaling 16% of the principals who responded to this question. While this is a low percentage, it is included in as sub-theme for a reason. Within many of the answers, participants referenced practices that could lead to higher student achievement. It would be inappropriate to conclude they were alluding specifically to student achievement, but the possibility exists. Therefore, I felt it was important to touch on this sub-theme.

Two survey responses and two interview responses referenced student achievement specifically (Table 8). Additionally, the following survey response seemed to allude to it, "...to improve and raise the bar for school improvement purposes." Below are the specific responses from principal answers.

**Table 8.** *Responses Referencing "Student Achievement"*

<b>Responses Referencing "Student Achievement"</b>
"to improve student achievement"
"allowing students to have higher understanding and achievement"
"improve student achievement"
"to improve and raise the bar for school improvement purposes"
"trying to promote student achievement and understanding"

The second sub theme under accountability was identifying ineffective teachers. In their responses, five principals referenced that the intent of this policy was to identify ineffective teachers. In the survey, three responses indicated a focus on this and two of the interview

respondents referenced this topic equaling a total 16% of principals answering the question. Table 9 shows the specific responses from participant answers.

**Table 9.** *Responses Referring to “Identifying Ineffective Teachers”*

<b>Responses referencing “Identifying Ineffective Teachers”</b>
“were attempting to pass legislation that would ensure unsatisfactory teachers were not evaluated as satisfactory”
“more effective method of documenting teachers who should leave the profession”
“make it so there were more teachers rated unsatisfactory”
“if there is poor teaching and learning there has to be some way to document that and then be able to use that either for improvement or consequences”
“trying to weed out ineffective teachers ”

#### **4.3.2 Effective evaluation tool (teacher growth)**

Another theme that emerged was the intended purpose of implementing a more effective tool for evaluation. As stated earlier, the responses indicated a focus on using the tool to help teachers grow. Their answers did not indicate a focus on professional development as a means for growth. In the survey responses, nine of the answers referenced this area. In the interviews, four participants made reference to this. Overall, 47% of the participants specifically referenced an effective tool based only on the survey responses. Table 10 shows the specific responses from participant answers.

**Table 10.** *Responses Referencing "Effective Evaluation Tool"*

<b>Responses Referencing "Effective Evaluation Tool"</b>
"help teachers & administrators know exactly which area/domain a teacher may need extra help in or is very successful"
"allowing the implementation on of a defined rubric to communicate what effective teaching practice consists of within all 4 Danielson domains"
research-based system of evaluation
"Better understanding of the importance with best practices for on-stage engagement and off-stage planning"
"a tool for encouraging teacher reflection and growth and a protocol for supporting conversations about teaching and learning"
"get the most out of teachers"
"trying to present the steps of what they should be trying to accomplish in each lesson that they teach"
"Trying to establish better teachers"
"Having the administrators go into classrooms and look for specific things"
"This actually gets administrators into classroom, looking at lessons. It starts conversations with staff members. I think policy makers wanted to ensure that teachers were reflecting and working on their practice. To an extent, the document does that, or at least starts conversations"
"I want to hope that it really was done to improve teaching and learning"
"I think their goal was to try to improve the instruction that our students are receiving"
"continue to grow teachers professionally"

### 4.3.3 Consistency

One theme that I was surprised to see was the theme of consistency. The answers appear to be referring to consistency among districts and buildings in how teachers are evaluated. This showed up in three survey responses and one interview. While this only represents 16% of the responses, I have included it because it was unexpected and referenced by several participants.

Table 11 shows the specific responses from participant answers.

**Table 11.** *Responses Referring to “Consistency”*

Responses Referring to “Consistency”
“To create a single, research-based system of evaluation”
“ensured all teachers were evaluated using a standardized method”
“Consistency”
“Something that is consistent across the state too would be one of the other motive to try to build some more consistent accountability across the state”

## 4.4 DATA COLLECTION AND ANALYSIS OF RQ2

*RQ2: How have principals been prepared to implement the Act 82 Educator Effectiveness policy in their schools?*

RQ2 was an effort to explore principals’ perceptions related to how they were prepared for implementation of the policy. This falls under the area of policy instruments (i.e. training) and capability (i.e. familiarity or prior experience with the 4 domains). Data collected in this

area came from survey and interview questions specifically focused on instruments and capability and was coded into two main categories: instruments, and capability. The survey results are presented first beginning with an overview of the findings followed by the data. The interview results are presented second.

#### **4.4.1 Survey analysis summary**

Overall, the survey results indicated all principals (n=27) were trained prior to implementation and all, except one, indicated that they were familiar with the Danielson four domains prior to implementation. The majority of principals (n=14, 52%) responded that prior to implementation of Act 82, they observed teachers using another evaluation tool based on the four domains. In addition, some of the principals (n=11, 41%) reported being observed as a teacher with an evaluation tool based on the same domains. This data suggests that the participants had some level of capability prior to implementation. All had participated in training (instrument) and over half had experience with a similar tool (capability).

As was mentioned, all participants had training. The majority (n=23, 92%) participated in training through their IU which is not a surprising result since this is a state policy, and the IUs role is to provide support and training for state initiatives. Most participants reported participating in training for 1-3 days (n=20, 72%). All 27 principals responded to the question asking about their level of preparedness. Overall, they reported feeling some level of preparedness for implementation with (n=23, 85%) describing themselves as either very prepared or prepared. The others (n=4, 15%) categorized themselves as somewhat prepared. No one indicated feeling not prepared.

Participants were asked what instruments were effective in helping them prepare for implementation. In the area of instruments, trainings provided by the IUs and districts appear to be perceived by principals as effective in preparing them. The data also suggests that an opportunity to work with the tool, and therefore build capability, was also effective in increasing their level of preparedness. Other instruments were found to be effective, but fewer principals utilized them. For purposes of defining “effective,” responses were considered to be indicating “effective” if principals responded in one of the following three categories: extremely effective, very effective, or moderately effective. Responses indicating “slightly effective” were not included.

Responses showed that participants found the IU trainings to be effective (n=20, 71%). District training had the same response with 71% (n=20, 71%) of responses indicating a level of effectiveness. The majority of participants did not participate in web-based training. Of those who responded to its level of effectiveness (n=9), 89% (n=8) reported it as being effective. There were 15 principals who responded to level of effectiveness for professional learning communities (PLC). Out of this group, 87% (n=13) rated it as effective. Independent learning activities was similar to IU and district trainings. Over half the participants (n=19, 68%) indicated it was effective. For those who had used a similar evaluation tool prior to implementation (n=11), 43% indicated that their prior experience was effective in preparing them. Finally, 72% (n=20) indicated that gaining experience with the evaluation tool was effective.

These results indicate a combination of instruments and capability building helping principals implement the evaluation tool. In the area of instruments, trainings provided by the IUs, and districts, appear to be perceived by principals as effective in preparing them. Other

methods of preparing were also found to be effective, but fewer principals indicated that they participated in them. The data also suggests that an opportunity to work with the tool and therefore build capability also was effective in increasing their level of preparedness. This is consistent with the literature covered in Chapter 2.

#### **4.4.1.1 Survey statistical analysis**

Various statistical tests were run to determine if there were any relationships between variables. The primary research variables included reports of feelings of preparedness, effectiveness of and duration of each training method (intermediate unit, district, web-based, PLC, independent learning). Additionally, perceptions of effectiveness of prior experience and use of the tool were examined. Overall, no relationships were found with the exception of one. A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in IU training effectiveness between the levels of preparedness. This result suggests that participants who felt they were somewhat prepared for implementation found the IU training more effective than participants who rated themselves as prepared. This result may have occurred because the IU trainings were designed as introductory trainings. For principals who already felt comfortable with the observation portion of the tool, introductory level trainings most likely did not add to their existing skill level. As a principal, I attended these trainings. Prior to implementation I had extensive training in another state with an evaluation tool based on the Danielson model. Therefore, I would have described myself as “prepared” and as a result did not gain much value from the trainings.



#### **4.4.1.2 Statistical testing**

Intellectus Statistics software was used for this section of the analysis. Intellectus Statistics is a software technology developed to efficiently and accurately conduct statistical analyses, interpret the output, and report the findings. The sample consisted of 28 participants; all but one respondent answered the majority of the survey questions. The primary research variables included reports of level of preparedness, perceptions of effectiveness of and duration of each training method (intermediate unit, district, web-based, PLC, independent learning). Additionally, perceptions of effectiveness of prior experience and use of the tool were examined.

#### **4.4.1.3 Testing results**

The initial plan was to conduct a Chi-square analysis. This was completed; however, due to insufficient cell sizes, the analysis could not be used. Further, examination of the descriptive data indicated that almost all participants had taken part in intermediate unit training, but the use of the other methods was more variable (between 18 and 54% of the sample reported that a given training method was “not applicable” to them, indicating the method was not used). For this reason, the inferential analyses focused on intermediate unit training effectiveness and duration, along with perceptions of preparedness, prior experience with the Danielson model, and use of the evaluation tool. Additionally, differences between the two counties on each of these variables was assessed.

Two tests were run to determine if there were differences in IU trainings by county. Specifically, these tests looked at duration and effectiveness. The findings did not indicate significant differences between the counties. A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in Intermediate Unit Training Duration between the counties. The Kruskal-Wallis test is a non-parametric alternative to the one-way ANOVA and

does not share the ANOVA's distributional assumptions (Conover & Iman, 1981). The results of the Kruskal-Wallis test were not significant,  $\chi^2(1) = 2.04$ ,  $p = .153$ , indicating that the mean rank of Intermediate Unit Training Duration was similar for each county. Table 12 presents the results of the Kruskal-Wallis rank sum test.

**Table 12.** *Kruskal-Wallis Rank Sum Test for Intermediate Unit Training Duration by County*

Variable	Mean Rank		$\chi^2$	df	p
	Beaver County	Washington County			
Intermediate Unit Training Duration	11.91	16.05	2.04	1	.153

A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in Intermediate Unit Training Effectiveness between the counties. The results of the Kruskal-Wallis test were not significant,  $\chi^2(1) = 3.75$ ,  $p = .053$ , indicating that the mean rank of Intermediate Unit Training Effectiveness was similar for each county. Table 13 presents the results of the Kruskal-Wallis rank sum test.

**Table 13.** *Kruskal-Wallis Rank Sum Test for Intermediate Unit Training Effectiveness by County*

Variable	Mean Rank		$\chi^2$	df	p
	Beaver County	Washington County			
Intermediate Unit Training Effectiveness	11.28	17.05	3.75	1	.053

Another test was run to determine if the variables of county and preparedness were independent. The results indicated that principal perceptions of their preparedness may not be dependent on which county they work in. A Chi-Square Test of Independence was conducted to examine whether County and Preparedness were independent. There were two levels in County: Beaver County and Washington County. There were three levels in Preparedness: Prepared, Somewhat Prepared, and Very Prepared.

Prior to conducting the analysis, the assumption of adequate cell size was assessed, which requires all cells to have expected values greater than zero and 80% of cells to have expected values of at least five (McHugh, 2013). All cells had expected values greater than zero, indicating the first condition was met. A total of 33.33% of the cells had expected frequencies of at least five, indicating the second condition was met. The results of the Chi-square test were not significant,  $\chi^2(2) = 2.44$ ,  $p = .295$ , suggesting that County and Preparedness could be independent of one another.

An interesting finding appeared in a test looking at principals' level of preparedness and how they rated the effectiveness of the IU training. A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in Intermediate Unit Training Effectiveness between the levels of Preparedness. The results of the Kruskal-Wallis test were significant,  $\chi^2(2) = 9.71$ ,  $p = .008$ , indicating that the mean rank of Intermediate Unit Training Effectiveness was significantly different between the levels of Preparedness. Table 14 presents the results of the Kruskal-Wallis Rank Sum Test.

**Table 14.** *Kruskal-Wallis Rank Sum Test for Intermediate Unit Training Effectiveness by Preparedness*

Variable	Mean Rank			$\chi^2$	df	p
	Prepared	Somewhat Prepared	Very Prepared			
Intermediate Unit Training Effectiveness	9.70	20.38	17.71	9.71	2	.008

Since the overall test was significant, pairwise comparisons were examined between each level of Preparedness. The results of the multiple comparisons indicated significant differences between Prepared-Somewhat Prepared. Table 15 presents the results of the pairwise comparisons.

**Table 15.** *Pairwise Comparisons for the Mean Ranks of Intermediate Unit Training Effectiveness by Levels of Preparedness*

<i>Comparison</i>	<i>Observed Difference</i>	<i>Critical Difference</i>
Prepared-Somewhat Prepared	10.68	10.30
Prepared-Very Prepared	8.01	8.38
Somewhat Prepared-Very Prepared	2.66	11.48

This data suggests that participants who felt they were somewhat prepared for implementation found the IU training more effective than participants who rated themselves as prepared. This finding may have occurred because the IU trainings were introductory in nature.

#### 4.4.2 Qualitative analysis—Interview

The following interview questions were used to gather data about principals' preparation for implementation as shown in Tables 16 and 17.

**Table 16.** *Interview Questions-Instruments*

How were you prepared, or what kind of training did you receive in order to successfully implement the new evaluation system?
Since the implementation of Act 82 in your district, have you had any additional professional development or support provided? Supports could include time to reflect with your colleagues, support provided by central administration, training at the intermediate unit, online training, etc.
Was there anything in your past experience, training or evaluation tools you worked with, that helped prepare you for implementing this evaluation system?
Did you feel you were adequately prepared?

**Table 17.** *Interview Questions-Capacity*

Based on what you remember, what was your experience like with teacher evaluation when you were a teacher?
During your time as an administrator, did you have an opportunity to evaluate teachers using a method other than what you are using now? If yes, can you please describe it and talk about how it compared to your current method of evaluation.
Did you feel you were adequately prepared?

Coding began by reviewing the transcripts with the first round focusing on references to instruments and capability. Specifically, in the area of instruments, I was looking for references to trainings and/or tools that were used to provide support in the implementation. In the area of

capability, I was looking for participant's prior experience with the evaluation method, which would indicate at least some knowledge and/or familiarity with it. Research indicates that the less of a gap in capability with what is being implemented, the more likely the implementation will go well.

The interview transcripts were reviewed and any reference to a policy instrument or capability was highlighted and copied into a separate document. Each reference was coded as instrument, capability or a combination of both. A second round of deductive coding occurred by reviewing the comments made about instruments and capability.

Analysis was done comparing participants' experiences in both areas with their feelings of being prepared. Specifically, I was looking for evidence that increases in capability, effective instruments and/or a combination of both would lead to higher levels of feeling prepared. This information was also analyzed to see if it suggested differences between support/training provided by the two I.U.s that serve these districts. This process was repeated when looking at capability, except analysis also included looking for variations and consistencies between individual participants. Throughout this process, notes were written to capture overall impressions, any data that was surprising, and reflections on analysis. This suggested an emerging pattern among the interview participants that capability grew during the implementation system leading to a higher, and possibly more effective, use of the teacher evaluation tool.

#### **4.5 INTERVIEW ANALYSIS SUMMARY**

All except one principal reported feeling prepared for the initial implementation of Act 82. Based on what the principals described, it appears that initial capability mattered to them in feeling prepared to implement a new policy which is consistent with what is reported in the literature. This was the most prevalent theme to emerge throughout the interviews. The principals described prior experiences with a similar evaluation tool as having an impact on them feeling prepared. Specifically, they were describing familiarity with the four domains of instruction that comprise 50% of the evaluation and are the basis for classroom observations.

It was interesting to note their comments about their colleagues, who did not have prior experience, or comments about experiencing more challenges with implementation. The one principal who reported not feeling prepared had the least number of years in administration. It is possible that she felt this way since she was still building capability as a building principal. Managing the implementation of a policy added another layer where she was required to build capability. This cannot be definitively concluded based on her comments. It can only be suggested as a possible cause. All of the principals referenced a growing level of comfort with the new evaluation tool as they continued to work with it. Their comments as a whole indicated that it is a tool that takes time to learn and the process is on-going.

In the area of instruments, the principals talked about the IU trainings and professional development offered by their districts. Their comments indicated that they found these trainings helpful. Their responses did not indicate any significant differences between the trainings offered by each of the IUs. This was not a surprise since it would be expected these trainings would be fairly standardized. They also indicated that the district sessions were helpful, but the nature of those varied since those were determined by the individual districts. One thing that did

come up, which was not anticipated, was discussion about software tools. Several principals mentioned using an online tool such as Pa-Etep and Towermetrix. These programs help manage the evaluation process and also contain training components. Their comments indicated that these tools were helpful in the implementation.

The finding that was most surprising was the emphasis on professional dialogue with colleagues. I was not anticipating this. As was mentioned in the paragraph on capability, the principals talked about gaining comfort with the tool as they continued to gain experience using it. Part of this process of gaining experience included professional dialogue. They mentioned multiple times talking with colleagues throughout the process and how helpful this was for them. They also talked about conversations with colleagues during trainings and initial implementation.

Following are descriptions of the interviews with each of the principals. Since principals were key players in this implementation, I felt it was important to share their perceptions of the process.

#### **4.5.1 Interview participant one**

Principal Miller is currently serving as a secondary principal in Washington County. In describing her experience with teacher evaluation, she indicated that she felt prepared and comfortable with the current evaluation tool. There were several factors that she identified as increasing her capability and familiarity with the current evaluation tool. The most significant was her previous, out-of-state experience.

Prior to working in her current district, Principal Miller completed a principal certification training program and served as an administrator in a different state. She identified both of those factors as being significant in her feeling prepared for Act 82 implementation. She



specifically referenced two specifics from her principal training program. The first was that it was an “on campus” program versus “online.” She felt this was important because as she stated, “I think that's the big difference I'm seeing with the interns that I'm getting here. They're doing online stuff and it's just not the same.”

The second factor was related to expectation. At the time, the state she was working in had a teacher evaluation that was very similar to the one we have now. She stated, “I learned it in my principal classes. It was expected that I knew it when I got to my job.” Principal Miller described the evaluation system being used in her state as similar to the one she using now. She described the process and stated they would “rank teachers on certain standard and provide feedback similar to what we do in Pennsylvania right now except without the teacher input.”

The data suggests that Principal Miller’s capability level, prior to implementation, had been developed based on her previous experience. The instrument that seemed to be most impactful was her principal certification program. It is interesting to note her comments about the importance of an on campus program. What was surprising during the interview was Principal Miller’s reflections on her colleagues’ experiences with Act 82 implementation. She explained that her colleagues had not had prior experience with the four domains that comprise the observation piece of the evaluation. She described their experience:

The other principals in my district struggled with providing that feedback to the teachers. They were used to just giving 20 straight across the board. Now that they were actually having to use education language, actually critiquing a teacher, providing open and honest feedback, it was hard for them. They really struggled.

She went on to explain that because of her previous experience, she found the implementation easier, and she felt she was more accepting of it, because she was “used to it.”

During the interview, we talked about Principal Miller's experiences with the implementation of Act 82, which occurred during her time as a principal in Pennsylvania. She shared that she received training from the Washington County Intermediate Unit (IU) which she indicated was a positive experience. "Going to the IU is best, training at IU helped me feel better prepared." I asked her about any on-going training since the initial implementation. Her answers indicated that she has not been involved in any additional training. She said she thought she might have gone to some additional trainings at the IU, but she couldn't remember.

Overall, Principal Miller perceived herself as being prepared for the implementation of Act 82. She credited her preparedness to her experiences during her out-of-state time. Specifically, she mentioned her principal preparation program (instrument/training), working with an evaluation system that was similar to what is being used currently (capability/familiarity) and trainings (instrument) provided by the IU. It is interesting to note that it is possible she has also increased the capability of her teachers for this implementation. She referenced providing training for them and being comfortable with doing so. Data suggests an effective implementation as she stated that:

They're using the language. They're talking with colleagues. That's a big thing. They never did before. They didn't share with each other but now they share. They talk about their evaluations. Okay, you've got a proficient. What did you do to get to proficient? That never happened. I think it's a good thing.

This suggests professional dialogue around effective instruction which aligns with the intended purpose of ensuring high quality teachers.

Principal Miller's responses highlight the impact prior experience with the tool can have on implementation. In her particular case, implementation was a fairly comfortable process

because she had extensive experience with the tool prior to implementation. In other words, the new tool did not represent a large departure in practice for her. It's interesting to note her comments about her colleagues. They tended to struggle due to the new evaluation tool being a larger departure from their previous experiences.

#### **4.5.2 Interview participant two**

Principal Jones is currently serving as a secondary principal in Beaver County. In describing her experience with teacher evaluation, she indicated that she did not initially feel that she was adequately prepared to use the evaluation tool. When asked if she felt prepared, she responded, "Oh no, absolutely not." Having now worked with the tool for several years, she stated that she does feel comfortable with it. Her answers suggested that her comfort increased as she gained more experience with it. She stated, "I'm comfortable with it. I've kind of honed the past two years, the process. Every time I do one, I think that I get better at it." Overall, her answers indicated that she participated in trainings which she found helpful, but it was gaining experience (e.g., increase in capability) that helped her feel proficient with the evaluation system. Following is a summary of her experiences in both areas of capability and instruments.

Principal Jones was evaluated as a teacher using the current evaluation system suggesting some level of capability since she had familiarity with the evaluation tool. Her experience occurred at the time that her district served as a pilot site. She describes her early experience as a teacher working with her principal, "I volunteered for the pilot. It was really working through that document. I was new to it. The principal at the time was very new to it. We had to work together to figure everything out." This suggests early capability building with the tool she

would later use to observe teachers. As an administrator, this is the only evaluation tool she has used to observe teachers.

Along with having familiarity with the evaluation tool, Principal Jones described trainings she participated in both as a teacher and an administrator:

I believe, as a teacher, we had some trainings on what the Danielson model looked like.

We looked at the rubric together, the framework. I also attended some sessions at the Beaver Valley Intermediate Unit as a teacher, which were very valuable.

When Principal Jones moved into the principal role, she contacted the IU to attend a training. She described this training as intense. She said there were only two other principals there so it was small, and during their time together they went through every part of the rubric. She went on to describe it as the standard training the IUs were doing at the time, but since there were only a few principals in attendance they were able to work through the material together and have discussion about it. Overall, she described it as “very beneficial.”

She did mention that they did not cover how to evaluate non-teaching professionals such as school nurses and guidance counselors. As a result, she described the process of evaluating them as “challenging.” She also suggested that other components in the training would have been helpful:

In addition, it would have been incredibly beneficial to either watch a videotaped teacher teaching a lesson and have everyone in that session actually do the evaluation components together to go through what does distinguished, need to improve, what does that look like? What is everyone's viewpoint on it? Or to watch an actual teacher do a lesson and do the same thing. There was no modeling of the actual process. That could have been very valuable.

This comparison between evaluating teaching professionals, for which she did receive training, non-teaching professionals, for which she did not receive training, suggests that the training she did receive was valuable in regards to her ability to implement Act 82.

Another factor that Principal Jones touched on was conversations with other administrators. She described these conversations as being helpful during her first year of working with the evaluation tool to observe teachers. Several times she specifically mentioned another principal in her district. She stated:

I've definitely talked to colleagues. That principal that I worked under when I was at the high school, I have a very close relationship with. I've spoken with him about the process, especially when I was new to it. The Superintendent at the time, I spoke with her about it just to ensure that I was on the right page.

She then referenced a second time the value of being able to talk through the process with the high school principal.

What was interesting to note from this interview was Principal Jones reflections on continued professional development. She described her initial trainings from when she became a principals, but she also referenced several times that additional training would be valuable. Her answers suggested that now that she has more familiarity with the evaluation tool, additional trainings would be beneficial. She stated:

I think like I said, some sort of actually modeling how to do the pre-observation meeting, watching a lesson, having a group of people sit down and fill the document out together. Talking about what does it look like in your building. That would have been very helpful. As well as like I said before, the ancillary staff members who aren't instructors. Any sort of refresher course would be helpful as well because sometimes you get stuck in a rut. If

there's something that maybe I should be doing that I'm not, but I don't know about it.

Any sort of just, "Let's check in and see how it's going training would be valuable."

Principal Jones's comments highlight the need for continued trainings beyond the initial, introductory trainings that were provided. Additionally, opportunities for professional dialogue should be built into these sessions. Principal Jones describes the one IU session that was very beneficial, and she mentioned the opportunities for conversations. In this particular case, the opportunity for conversation was not planned but rather the result of low participation in this particular session. Moving forward, these types of interactions should be intentionally built into the trainings.

#### **4.5.3 Interview participant three**

Principal Smith is currently serving as an elementary principal in Beaver County. In describing her experience with teacher evaluation, she indicated that when Act 82 was implemented she felt she had a good understanding of the domains and components that comprise the observation section of the evaluation. Similar to Principal Jones, she felt that she became better over time as she gained more experience using the tool. She described her early experiences:

It took me time implementing it with staff to really see connections with how to collect some of the data in classrooms as far as ... Like making a statement that most of the kids are raising their hand and participating. Being more specific about what percentage of kids or if it's the same kids who are raising their hands each time. So it took me some time to fine tune my observations and the type of information I was recording in it.

Prior to becoming an administrator, Principal Smith was evaluated under the previous system which utilized the same domain areas for evaluations so she felt she had some familiarity

with the new tool. The main difference she identified is how in-depth the new tool goes into the instructional domains. She described her experience of being observed as a teacher as positive, but said, “it wasn’t really a formal process.”

In the area of instruments, Principal Smith indicated two that seemed to provide the most support. The first was engaging in walk-throughs with other administrators and discussing what they observed. A walk-through is a more informal version of a classroom observation. It is typically short in duration, and is not written up as a formal observation. Principal Smith stated: “The best part I found for me was doing some walkthroughs and observations with colleagues.” She specifically mentioned doing this with a principal colleague who worked at a different level. She said this principal noticed things in the classroom that Principal Smith would not have typically noticed. She described this experience as “enlightening because things stood out to her that I just kind of take for granted and didn’t pick up on. That to me enriched the process of learning how to identify the components in the classroom settings.”

Another instrument Principal Smith referenced throughout the interview was the online evaluation program Pa-Etep which her district utilizes to manage the evaluation process. Pa-Etep manages the scheduling of meetings and observations, provides information on the components, houses all of the evaluation data and has embedded training videos. She described an initial training on Pa-Etep and then continued on describing her experience with it. At one point, she mentioned that, “I’m able to get a holistic picture of each of our staff and capture those observation and data using PA-ETEP.”

As a follow up, I asked her if she found that using PA-ETEP to manage all of the data and information had helped her be able to focus more on the observation and instruction. Her response was:

I do, yeah, I do feel that way. Not only is it managing all the information, I also see it as improving communications with teachers. While it is an online format, we are able to take time communicating and I do have comments from teachers and we will communicate back and forth using the field and PA-ETEP which makes it easier if I'm not able to catch them during their planning period or right after school or something. I feel it facilitates conversation and also holds us accountable too. It's easy to get caught up in other things and not follow through on the post observation conversation. I feel that having it managed through a system that gives sending reminders and facilitating that communication back and forth that builds in another layer of accountability both in the teachers and for myself.

There were several other instruments Principal Smith mentioned. These were other methods of training she received during the implementation. She briefly mentioned what she thought might have been a one day session at the IU. She also engaged in independent learning by reading articles and book excerpts on the Danielson model and discussions with other administrators during district meetings. It was interesting to note, that she stated she has not had any additional formal trainings since the implementation of the evaluation system.

Principal Smith's responses highlight several areas that are covered in the Ch. 5 recommendations. The first one is the opportunity to observe with another administrator. This provides an opportunity for an embedded professional growth opportunity. Administrators have an opportunity to discuss what they observed and how it fits into the Danielson framework. It also creates a situation where multiple people observe a teacher which is an effective evaluation practice covered in the literature. The second area is on-going, independent learning



opportunities. In Ch. 5, recommendations are made regarding how districts can support this type of professional growth.

#### **4.5.4 Interview participant four**

Principal McWilliams is a secondary principal serving in Washington County. With over twenty five years in education, he has the most experience of all the interview participants. His last experience of being observed as a teacher was over 15 years ago. He described that experience as being completely different from how teachers are observed and evaluated now. He described it as “very simple” with the principal stopping in the room and no follow up feedback. He said he would get a form in his mailbox and be rated satisfactory at the end of the year. Yet, he indicated that experiences he had while serving as an administrator helped develop his capability so that he felt comfortable implementing Act 82. He described his experience with the Danielson model prior to implementation:

We generated our own district-wide teacher evaluation tool that we came up with as an administrative team. Then from there, doing research I found the Danielson Model a long time ago back in probably 2006. I started using that and created our own model using the Danielson Model before the Act 82 came out. I was familiar with the Danielson Model prior to Pennsylvania having us use the teacher effectiveness model.

When asked if he felt prepared, he responded:

We always had the 4 domains, the planning, the instruction and the professionalism. It didn't take very long to see what the state wanted us to look at as far as the teacher model I thought we were adequately prepared to do that type of a supervision tool.

His answers indicated that through prior experiences in his district, he had developed his capability prior to implementation leading to a level of comfort with the new process. Similar to the previous principals, he indicated that the more he used the system the better he was at it. He referenced it becoming more familiar and stated “It took time to learn the system.” This suggested an emerging pattern among the interview participants that capability grew during the implementation system leading to a higher, and possibly more effective, use of the teacher evaluation tool.

Principal McWilliams referenced several instruments during our conversation. He shared that his initial formal training was viewing the “the Danielson Effectiveness Model video modules” which showed a pre-observation conference, classroom observation and post-conferences. He described this as helpful, “It was helpful to watch those modules to see the questions that you should ask a teacher prior to observing, looking at the planning of the different domains.”

He also shared that he attended a few trainings at the IU, but his comments did not indicate that those were particularly helpful for him due to his prior knowledge of the domains, but he did indicate that they were helpful for his colleagues, “I had experience using the 4 domains prior to, but the other administrators in our district seemed to pick up on it pretty quickly.” When asked about any additional trainings, he said he had engaged in all of the ones I mentioned which were time to reflect with colleagues, support provided by central administration, training at the intermediate unit, online training or any independent learning. He went on to elaborate that he is completing his dissertation on the walk-through process and that he has trained other administrators on this.

The final instrument he mentioned was his district's use of Towermetrix. This is an online software similar to Pa-Etep which manages the entire evaluation process. Principal McWilliams referenced teachers answering pre-observation questions in the program and use of all of the observation data that is housed in there throughout the observation process.

Similar to the other respondents, Principal McWilliams indicated that his prior experiences were helpful in preparing him for implementation. He also referenced several times independent learning activities he has engaged in. This highlights a way districts can support principal growth without significant financial investments. This topic is covered further in Ch. 5.

#### **4.5.5 Interview participant five**

Principal Jackson currently serves as an elementary principal in Washington County. When asked about her level of preparedness with the evaluation tool, she indicated that she was prepared. The majority of her comments, as she described her level of preparedness, fell into the area of capability. She shared the following, "I do feel like I was adequately prepared but I feel like without the prior experience of being evaluated under that tool I don't know if I would have been as prepared."

Her experience with the current evaluation model began during her time as a teacher. She explained that in her last 6 years of teaching her district utilized an evaluation tool that included a rubric based on the four Danielson domains of effective instruction. She also shared that she felt having been observed using this tool helped her gain credibility with her teachers when she became an administrator. She stated:

I was going to say that's actually been helpful for me as an administrator in helping people to navigate that system a little bit. I've had experience as an educator using that

system so I feel like when I was giving teachers information about it being able to say that I lived under something like that helps with your credibility a little bit.

Prior to serving in her current role, Principal Jackson served as an assistant principal. During this time, she evaluated teachers with the current evaluation tool. As she described this, she shared that they focused on conducting walk-throughs. Each month they identified a different area to focus on for the walk-throughs, and they completed a narrative based on the four domains. After explaining this process, she returned to her earlier focus about familiarity with the tool:

I feel like until you're doing it you have to be familiar with the evaluation tool and that rubric. I think until you've had experience with that rubric it's hard because there's so many different areas that you're looking at. I felt prepared because like I said, when I was a teacher we were scored on a Danielson rubric and it was very very similar. In fact, very little has actually changed from that so I felt prepared to evaluate people because I was very familiar with that tool and comfortable with it.

Principal Jackson's final remarks in the area of capability referenced her colleagues' experiences with implementation. She described their experience with it as possibly being more challenging, because they lacked the prior experience that she had with the tool. She explained:

I do feel like in listening to colleagues and things, them talking, the issue seemed that some people just weren't familiar with the tool. Then you were trying to figure out where to put your comments about different areas and where did things fit.

Principal Jackson touched on the topic of instruments, but her overall comments did not indicate that she felt they impacted her experience as much as her capability. She stated that she attended trainings at the IU and still attends ones when they are offered. She referenced some

additional “mini trainings” that were offered by her district. She also shared that the administrative team continues to have on-going discussions. She described her district experience as follows:

We definitely through our administration and leadership meetings have had discussions and we've looked at our observation forms that we used to fill out. We keep revising and editing to make it more user friendly and fit better with that tool. I do think talking with colleagues helped, I've gone to continued trainings.

Principal Jackson added an interesting piece that had not previously come up in interviews. She referenced having the right mindset about the evaluation tool. She described it as follows:

I do feel like you have to have the right mindset about the tool. I think it's very overwhelming for teachers and for administrators if people are trying to be distinguished in every single area so I've tried to talk to teachers about, it's a tool for growth and you should work on having goal areas within the tool that you're trying to improve in and if you have those goal areas set and put your focus on that you show improvement in certain areas. Then the following year you pick new areas, you keep those areas you worked on the previous year high and then the following year you pick new areas of growth to work on improving and then over time you really see your improvement as an education as a whole. We see your improvement as an educator as a whole so we've done a lot of work in my building around that doing it that way.

Other participants, in both survey and interview responses, referenced talking to teachers about how you “live in proficient and visit distinguished” which is a quote often used by Charlotte Danielson. Interestingly, Principal Jackson’s comments framed it in a way that had not

previously appeared in participant comments. This speaks to a value system which can impact implementation of policy as was discussed in Chapter 2.

Principal Jackson's comments were similar to the other principals in regards to the value of prior experience with the tool being beneficial during implementation. In addition, she added that she felt she had more credibility with the teachers, because she had also been observed as a teacher under this framework. A unique focus she brought up was mindset. She was the only principal to specifically touch on the idea of helping teachers develop a positive mindset about the evaluation tool.

#### **4.5.6 Interview participant six**

Principal Kyle is a secondary principal in Washington County. When asked about his level of preparedness with the evaluation tool, he indicated that he felt confident about using the new tool. He explained:

I felt pretty confident, because when I started, the guy that was the principal at the time, him and I did two walkthroughs together. Not two walkthroughs, two observations prior to me doing one on my own, and then I'd been to the IU for the different trainings at that point, so I felt pretty confident when I went in to do my observation on my own.

Similar to Principal Jones, Principal Kyle referenced a fellow administrator who was instrumental in helping him feel prepared initially. He described inservice days during implementation phase which occurred while he was still in a teaching position. These trainings were provided in his district by the superintendent. He explained that he had a lot of "side conversations" with the superintendent, because she was a family friend. He indicated that those conversations and inservice days were helpful in his early capability building with the new

evaluation. Like Principal Jones, he described a relationship with his former superintendent in which he was comfortable asking questions:

The biggest thing I would say where I really learned a lot was speaking with the superintendent ...because she was at the forefront with all of this stuff going on. The personal relationship that I've had has helped me ask professional questions to her, and I really feel that that was one of the biggest things that helped me.

Principal Kyle said that his early observation experiences as a teacher were nothing like the current evaluation system. The experience just described occurred later in his teaching career as Act 82 was being implemented. Therefore, his capability was developed prior to entering administration.

Principal Kyle mentioned two instruments from his time as a principal. The first was trainings at the IU. He described attending several over time and said he picked up something from each one. This aligns with the research that recommends learning be on-going. It may have been that as his capability developed, he was able to gain different information from the trainings. He also shared that his district uses Pa-Etep to manage the evaluation process. He did not indicate if using this system had helped him develop his skills in this area.

Principal Kyle's comments echo the other principals in highlighting the importance of professional dialogue and collaboration. He also touches on the idea that trainings need to be on-going. Specifically, he mentioned that he continued to attend trainings and picked up new information at each one. This supports the idea of the interactive nature of the relationship between instruments and capability. As capability increases, principals will continue to gain new insights by having access to continued professional growth opportunities.

## **4.6 DATA COLLECTION AND ANALYSIS OF RQ3**

*RQ3: How have principals implemented the Act 82 Educator Effectiveness policy in their buildings?*

RQ3 was intended to provide focus on how Act 82 has been implemented in schools. This falls into the category of environment. Data was analyzed in two main areas: effective evaluation practices and professional development. A priori coding was used to divide data into these two main categories. In the evaluation practices, I looked for examples of effective practices i.e. effective tool, trained observers, multiple observers for each teacher and use of student data. In the category of professional development, I looked for evidence of embedded, on-going professional development, and evidence that data from teacher evaluations was taken into consideration when planning for PD. Data was categorized into these two areas: PD is embedded and on-going and data from teacher evaluations is considered when planning for PD. I also looked for any unanticipated responses that fell outside of these categories. Once coding was complete, the data was analyzed to look for evidence that effective supervision practices and professional development were or were not occurring and variations that existed between participants and districts.

### **4.6.1 Effective practices**

Data for this category was collected from participant answers to the questions shown in Tables 18 and 19 on the following pages.



**Table 18.** *Survey Questions-Effective Practices*

Have you ever participated in training on Educator Effectiveness?
Please indicate the type of training you had on Educator Effectiveness and approximate length of time. If you did not participate in training, please mark “not applicable.” (matrix answer)
Based on your experience in your current building, indicate your level of agreement with the following statement, “The Educator Effectiveness evaluation model has helped ensure there is an effective teacher in every classroom.”
How many administrators typically observe each teacher in a school year?
Is there anything else you would like to share about your experience with the Educator Effectiveness evaluation model?

**Table 19.** *Interview Questions-Effective Practices*

How were you prepared or what kind of training did you receive in order to successfully implement the new evaluation system?
Do you feel the new evaluation method has had an impact on the quality of instruction?
In general, what’s been the reaction of your teachers to having student data be a part of their evaluation?
Are you the only one who observes teachers in your building or are they observed by other administrators?
Overall, how well do you think the teachers in your building understand the four domains that comprise the “observation/practice” portion of the rubric?
How many administrators typically observe each teacher in a school year?
Has the new evaluation system impacted your ability to remove ineffective teachers?
Have there been any unintended consequences, positive or negative, that have occurred as a result of the current evaluation method?
Is there anything else that you would like to share with me about your experience with teacher evaluation?

#### 4.6.2 Effective tool—Survey responses

The first component of effective evaluation is the evaluation tool. Teachers and principals need to understand the expectations of the evaluation tool and be able to translate those expectations into practice (Bill & Melinda Gates Foundation [Gates Foundation], 2012; Darling-Hammond, 2013). In the survey, participants were asked to indicate their level of agreement with the statement, “The Educator Effectiveness evaluation model has helped ensure there is an effective teacher in every classroom.” The responses indicated that many of the principals neither agree nor disagree (40%) while only about 25% stated agreement with this statement as shown in Table 20.

**Table 20.** *Educator Effectiveness*

<b>Participant Response</b>	<b>Number</b>	<b>Percentage</b>
Strongly Agree	0	0
Agree	7	25.9
Neither Agree or Disagree	11	40.7
Disagree	1	3.7
Strongly Disagree	0	0

The participants were also asked about their teachers’ level of understanding in the survey question, “Overall, how well do you think the teachers in your building understand the four domains that comprise the ‘observation/practice’ portion of the rubric?” The majority indicated an understanding ranging from moderately to very well as shown in Table 21.

**Table 21.** *Four Domains*

<b>Participant Response</b>	<b>Number</b>	<b>Percentage</b>
Extremely Well	0	0
Very Well	11	40.7
Moderately Well	13	48.1
Somewhat Well	2	7.4
Not well at all	1	3.7

At the end of the survey, participants were asked if there was anything else they would like to share. The responses were revealing. There was inconsistency in their answers regarding how effective this tool is for evaluation. Comments ranged from negative, stating that the process is too time consuming, and it's now more difficult to remove ineffective teachers as shown in Table 22 on the following page. Others were positive stating a greater focus on collaboration with teachers and professional growth. This is consistent with what the literature states about it being difficult to implement education policy in a consistent manner.

**Table 22. Sample Survey Comments**

<b>Sample Survey Comments</b>
“The new system is very frustrating when you are working in a high performing school. It is nearly impossible to "rate out" a poor teacher because the school level results count for so much of their evaluation (especially if they do not have a PVAAS score). The rubric defines the failing level of performance so poorly that it is nearly impossible to not reach the needs improvement level or proficient level in certain components.”
“I feel the Educator Effectiveness evaluation model has enabled me to focus more on providing feedback to teachers rather than a one and done evaluation that I personally as a teacher was so accustom to concerning classroom observations. I try to emphasize with my staff that the purpose of the Educator Effectiveness evaluation model is to provide feedback and professional growth. I encourage teachers to try new things in the classroom during this process and not be fearful if something does not go as planned.”
“The amount of time this process takes per teacher is ridiculous--it's just more paperwork. In my opinion, walk-throughs are more effective than the current process. Any teacher can put on the dog and pony show when they know date and time that I am coming to observe them...”
I am more engaged in paperwork using the new tool than ever before. This responsibility takes away from my ability to engage with students and parents and to work on new initiative in other areas.
I hope the state ends it. It is a challenging process that punishes administrators rather than improving instruction. In the past, it was easier to rate teachers as unsatisfactory and implement improvement plans. I believe the Educator Effectiveness model has made the observation process more focused on timelines than on effective supervision models.
“I don't prefer it. Effort is not worth the reward.”
“ It is clear to me that the legislature and PDE has taken a model of evaluation far beyond what was intended by Danielson.”
“Lengthy and very time consuming.”
“The collaboration between the administrator and teacher allows for one-on-one dialogue concerning professional growth.”

#### **4.6.3 Effective tool—Interview responses**

Overall, the interview participants reported a positive experiences with the new evaluation tool with the exception of one. This principal reported positive outcomes but was generally more critical. Interestingly, the principals reported a generally favorable response to the tool from the teachers with the exception of the aforementioned principal who described her teachers' experiences in more negative terms.

The theme that was most prevalent throughout all of the interviews was the idea that the new model has increased professional dialogue, collaboration and reflection. Every participant touched on this idea. The participants reported more discussions focused on instruction with their teachers as well as teachers dialoguing with each other about effective practices. Some of comments provided are listed below in Table 23.

**Table 23.** *Professional Dialogue and Collaboration Comments*

Professional Dialogue & Collaboration Comments
<p>“They're talking with colleagues. That's a big thing. They never did before. They didn't share with each other but now they share. They talk about their evaluations. Okay, you've got a proficient. What did you do to get to proficient? That never happened. I think it's a good thing.”</p>
<p>“It's a great piece to have conversations and look at areas of strength and areas of growth.”</p>
<p>“ I think that it has opened up other conversations which have improved instruction.”</p>
<p>“I really have seen it as a way of helping me be a better facilitator of connecting teachers and helping them develop professionally things that they have identified as goals in addition to things that I observe and want to support in their classroom.”</p>
<p>“I'm able to provide specific feedback I'm able to offer insight because I see it in other classrooms. I can shared what I've learned or seen in other classrooms with teachers.”</p>
<p>“I think it's getting teachers and educators to talk about teaching and helping teachers with their craft more so than some of the previous supervision tools. I do think it's helping teachers become better teachers if it's done correctly.”</p>
<p>“Over the last two years working as a team with different teachers in the building to improve so that you're collaborating and using each other's strength areas to improve overall. “</p>
<p>“I think it's opened up some areas of discussion for people. I think it's positively given people a way to pick focus areas and set goals for themselves where maybe that wasn't, when you were doing one observation over the course of the year and then you turning that in and that was the end of it, this gives you a way to keep that dialogue open.”</p>
<p>“I would say yes, it has impacted the instruction in various different ways, because I feel when we go in there, with the way that we're using the system, we're getting more information and the teachers are getting more feedback, and they're getting it quicker, whether it be through an observation or a walkthrough. We really feel that it has helped our instruction that we've seen in the classrooms.”</p>
<p>“...we're getting into the classrooms probably more than we had been in the past, and they've been able to have more feedback with administration, and we've allowed them to have time to collaborate with one another. “</p>

The theme of professional reflection also came up throughout the interviews. When talking about this idea, the participants were referring to teachers using the four domains to reflect on their own practices. While comments were generally positive, the participants did indicate that teachers were inconsistent in this area. As can be seen in the comments below in Table 24 some teachers embraced the new tool to help them become more reflective practitioners while others appear to simply comply with what they are required to do for the evaluation process.

**Table 24.** *Professional Reflection Comments*

<b>Professional Reflection Comments</b>
“They understand that they're not perfect and that I'm not out to get them. I am just trying to help them become a better teacher. They've gotten used to part of the evaluation; we actually make them write professional goals. Over the years their professional goals have become more reflective.”
“I think them having to reflect on themselves before they meet with me, I think it's having them think deeper. “
“I still get a handful that will still put proficient all the way down. They don't want to think about it.”
“I think that the tool is very beneficial for being reflective on practice. “
“There's a value of having that time to collaborate and see each other doing things in the classroom and a desire to want to learn from each other and share ideas that to me it is just really encouraging an education.”
“I see different levels of engagement with that from teachers who are just very concise in their thinking as far as responding to the prompt. I see teacher who write several paragraphs and are really giving me deep insight into how they are thinking and how they are working through some of the dilemmas and how difficulties that they encounter.”
“I have heard positive feedback about the prompts and how it gets them to think a little bit differently.”
“ I do feel like teachers that have been reflective and have chosen areas and really focused on those areas I have seen a lot of growth and improvement in them.”
“I have had a couple of teachers say I'm happy with whatever you rate me as knowing that they'll be proficient. They haven't had any interaction that would show them that they would be needs improvement or basic. It's interesting because I think there's a group of people who are happy to be status quo and they feel like they're coming in and they're doing their work. It's enough.”

The interview participants were asked if they thought the evaluation tool has had an impact on instruction. Four of them answered that yes it did. One indicated that it was a start and the other said not necessarily. Of the four who indicated a positive impact, they tended to



attribute the impact to the professional dialogue, collaboration and reflection that was covered in the previous section. The one variation was the participant who focused more on what her district is doing with the differentiated supervision model although it includes the professional collaboration elements. She shared:

One of the thing that comes to mind would be the differentiated supervision options. I traditionally have a lot of teachers in my building who want to get out and see others teaching in their classrooms. There are things they are working on specifically where they'd like to improve in their own practice. I feel that my ability to get out and observe is in walkthroughs and observations I can connect teachers with others who are doing things specific to what they want to learn about. I really have seen it as a way of helping me be a better facilitator of connecting teachers and helping them develop professionally things that they have identified as goals in addition to things that I observe and want to support in their classroom.

One participant answered that for some teachers the new evaluation tool has had an impact. He then continued on to describe elements of professional dialogue, collaboration and reflection. The final participant said she did “not necessarily” think it has impacted instruction but did go on to add, “I think that it has opened up other conversations which have improved instruction.” The concern she shared was how “broad” the rubric is for the four domains. She stated that she found the whole thing “overwhelming”, especially for new teachers. She provided the following example of this, “I have a new teacher this year, who I couldn't even get through the document with her, because she had made improvements and she was really taking it poorly. I had to stop the meeting.”

A negative perception that was brought up in several of the interviews was the amount of time and paperwork involved with the process. This was consistent with what was reported in the survey. While it was brought up as a concern, there was not a heavy focus on it. In other words, participants did not give the indication that it was so cumbersome that it detracted significantly from the positive benefits. Although it should be kept in mind that those who opted to be interviewed may be principals who generally viewed the experience as more positive when compared to their colleagues. This will be discussed further in Chapter 5.

Interestingly, four of the interview participants talked about use of a software tool (instrument) to manage the process. Three of them mentioned commercially produced products and one referenced an online tool that had been developed by a colleague in his district. Their comments suggested a more positive experience with managing the paperwork and scheduling that is inherent in the implementation of this policy. This aligns with research in that having an effective instrument can help with implementation of a policy.

#### **4.6.4 Trained observers—Survey responses**

The second component is trained observers. To use any evaluation tool correctly, the observers need to complete training and demonstrate competency (Gates Foundation, 2012; Darling-Hammond, 2013). All of the participants were asked if they had some type of training on the new evaluation system. Everyone responded “yes.” They were then asked about what kind of training and length of time. All except 1 indicated that they had training through their IU.

The length of time for the majority was 1-3 days of training. This is to be expected since this was a state initiative. The majority of participants indicated they had some type of training in their districts, but there was little consistency in length of time. Very few of the

trainings appear to be on-going. Participants who indicated on-going training responded that this was through some type of independent learning. Table 25 shows the type of training and duration participants engage in.

**Table 25.** *Training and Duration*

Question	Not Appl icabl e		1-2 hours		Half Day		1 Day		2-3 Days		5 Days or More		Training is on- going		Total
Training provided by the Intermediate Unit	3.70 %	1	3.70%	1	7.41%	2	37.04%	10	37.04%	10	3.70%	1	7.41%	2	27
Training provided by the school district	26.9 2%	7	19.23 %	5	7.69%	2	11.54%	3	19.23%	5	11.54 %	3	3.85%	1	26
Participa tion in a professio nal learning commun ity in my district or building	46.1 5%	12	15.38 %	4	3.85%	1	11.54%	3	15.38%	4	3.85%	1	3.85%	1	26
Web Based Training (e.g. Teachsca pe)	61.5 4%	16	7.69%	2	3.85%	1	3.85%	1	7.69%	2	11.54 %	3	3.85%	1	26
I have engaged in my own indepen dent learning	19.2 3%	5	0.00%	0	19.23%	5	0.00%	0	15.38%	4	15.38 %	4	30.77%	8	26

#### **4.6.5 Trained observers—Interview responses**

All of the interview participants engaged in multiple trainings prior to, during and after implementation. Every participant spoke about attending training at their Intermediate Unit. Overall, they found the trainings helpful. There were some variations in their comments. One principal stated that the IU trainings were helpful, but it was her prior principal certification program and experience that truly prepared her. She also went on to share that the other principal in her district, who did not have prior experience with a similar tool, really struggled with implementation despite attending the IU trainings.

Another principal said she found the IU trainings to be very valuable, but it is important to note her particular experience. She shared that there were only three people in the training so they sat with the trainer going through the materials in an in-depth manner. This was slightly different than the typical large group training. Another principal said the IU trainings were helpful and that he was able to learn something new in each one.

Four of the six participants said their districts provided some sort of training during implementation. Two of them explained that at least part of their district training was actually training on how to use the Pa-Etep software to manage evaluations. One went on to add that he has continued to access the webinars that Pa-Etep offers in their professional development library. Two participants said they have done some additional professional reading on the Danielson domains of effective instruction.

Overall, the participants indicated they found value in all of the different types of trainings they engaged in, but considering their answers holistically it must be noted that gaining experience with the evaluation tool seemed to be the most valuable. The data suggests that having continued training as the principals gain experience with the tool may be helpful in

continuing to develop capacity. As one participant stated, “Not one of them really sticks out as being better than the others. I think every one of them that I went to, I picked up something that has helped me along in the process.”

#### **4.6.6 Multiple observers—Survey responses**

The third component, multiple observations by different observers, focuses on ensuring accuracy of observations. A single observation may be adequate for the purpose of providing feedback, but when high stakes evaluation is involved multiple observations need to be completed to ensure an accurate picture of how a teacher is performing. The observations should also be conducted by more than one person. For example, if the principal is the primary observer, it is recommended that other individuals also complete observations. Other individuals might include central office administration, expert teachers and/or instructional coaches (Darling-Hammond, 2013; Gates Foundation, 2012).

In the survey, participants were asked how many administrators typically observe each teacher in a year. The majority (59%) responded that teachers are observed by only one administrator. Another 26% responded that teachers are observed by more than one administrator and 15% indicated teachers are observed by more than one administrator but only if the teacher is struggling. This data suggests that in most cases teachers are not observed by multiple administrators.

#### **4.6.7 Multiple observers—Interview responses**

Generally, the interview participants indicated that they are the only ones who observe their teachers. Two of the principals answered that they are the only person who evaluates their teachers. Three of the principals answered that they are the only person who evaluates unless the teacher is struggling. Then another administrator may observe that teacher, but general practice is only they evaluate. Only one principal indicated that his teachers are observed by other administrators, “They're observed by other administrators as well. There's an assistant principal that does it, and our assistant to the superintendent does a couple as well.”

#### **4.6.8 Evaluation tied to student data**

The final component is considering observation data in relation to student data. It is better to consider observation data in conjunction with student performance on various assessments, including state achievement tests. When considering achievement testing, it is not necessarily only about overall achievement, but taking into consideration students' starting points to determine growth made during the school year (Gates Foundation, 2012; Darling-Hammond, 2013). In the Educator Effectiveness model, student data is accounted for in the final rating. Therefore, this study did not look at if student data is included but rather how the inclusion of data has been perceived by the teachers. Interview participants were asked, “In general, what's been the reaction of your teachers to having student data be a part of their evaluation?”

The participants consistently reported teachers having mixed feelings about the student data piece. Some of the words they used to describe teachers' responses were excited, depressed, happy, upset, and nervous. Two themes emerged from their answers: inequity and control. The

first, inequity, refers to the perceived inequity of different accountability standards. For example, a seventh grade math teacher would have her student's data on the math PSSA and the building level score (SPP) count for her evaluation. A seventh grade Social Studies teacher would only have the SPP count since this subject area is not a tested area. One participant shared:

I think that some teachers feel that they're being, like a math teacher, they have more pressure on them than someone teaching social studies. The data, the 3-year data coming out now is not fair to them compared to other teachers that teach different subjects in the building.

An unexpected theme was control. This refers to the level of control teachers perceive having on the outcome of their data. The higher level of control they felt they had the less negative their feelings were about having data included. I specifically say "less negative" because the data indicated that teachers generally did not have positive feelings about this piece. Principals described teachers feeling they did not have control over the outcome. One example was a primary building. When an elementary building serves only grades K-2, their SPP is based on the intermediate elementary building's SPP scores. This has led to frustration because once the students leave their building, the teachers do not feel they have control over their outcomes yet they are held accountable for it.

Where teachers appear to feel more in control is with their Student Learning Objectives (SLO). This is a goal they create, and they set the ranges for proficient, distinguished, etc. This is done in cooperation with their principal. Principals indicated that teachers generally felt better about this part of the evaluation, because they have more direct control in setting the parameters and influencing the outcome.

During the interviews, principals were asked about how they talk with teachers about data. Their answers varied. Two principals described a structured process for data team meetings. One principal described meetings at the beginning of the year to review the standardized test results from the previous year and then teachers discussing data in department meetings throughout the year. She indicated this was an area she needed to work on. The other principals described more informal processes such as talking with teachers about their data during evaluation conferences and/or faculty meetings. There was also inconsistency in which data principals focused on. The data included a mix of standardized test scores, common assessments, PVAAS (student growth data) and universal screeners.

#### **4.6.9 Professional development**

The goal of Pennsylvania's Educator Effectiveness model is to reform how school professionals are evaluated as well as professional development (Pennsylvania Department of Education, 2014). This focus on training and development of teachers, based on evaluations, is supported by the literature. Evaluation alone does not necessarily improve teaching practice, but professional development based on feedback from evaluation can improve teacher effectiveness (Darling-Hammond, 2013). "Evaluations should trigger continuous goal-setting for areas teachers want to work on, specific professional development supports and coaching, and opportunities to share expertise, as part of recognizing teachers' strengths and needs"(Darling-Hammond, 2013, p. 99).

There are practices considered to be effective when delivering professional development that is intended to change instructional practices and impact student achievement. One is duration. Professional development (PD) should extend over time allowing teachers to go in



depth with the topic and it needs to include plans for ongoing follow-up (Garet et al., 2001; Porter et al., 2000). Another component is teachers having some choice in the topic (Porter et al., 2000) so it will be specific to the content the teacher is teaching. Job-embedded learning is the practice of educators sharing information, formally or informally, about what works and does not work, trying new practices and reflecting together (Zepeda, 2012). The critical piece here is educators having a opportunities for active collaboration (Garet et al., 2001; Porter et al., 2000). This type of PD provides educators an opportunity to engage in the transformational learning process by discussing and working through common issues.

In the survey, principals were asked, “When planning for professional development, does your district take into consideration feedback from teacher evaluations?” In the interviews, principals were asked, “Can you share with me how your district goes about the process of planning for professional development?” In the survey, principals were given three options to select from: yes, no or not sure. The majority (63%) indicated that their districts did take into consideration teacher evaluations when planning for professional development. A smaller percentage (33%) responded “no” and 1 principal responded “not sure.” This data suggests that the policy is being implemented as intended in most districts. The interview responses provided additional insight.

In the interviews, principals were asked how their districts go about planning for professional development. Their answers indicated evidence of the practices described in the research, but variations in their responses existed in all areas. The most prominent theme was using teacher feedback to determine priority areas. This aligns with the literature in allowing teachers to have choice.

All but one principal indicated that their districts take into account teacher feedback when deciding on professional development topic areas. Teachers have an opportunity to input what areas they feel they need to develop in or areas of interest. For example, one principal shared: “Each of our buildings has an instructional cabinet made of teacher leaders that are representative of the grade level or the content areas. That is a system that we can get feedback from teachers about areas of interest. I use my observations and conversations with teachers about things that they would like to work on or say that our areas of interest to them to determine if there are things that are areas of focus or if there's pockets of interest within our building so we can define some professional development or even opportunities for teachers to collaborate.

Another principal responded:

The teachers at the end of the year do a climate survey, and in that climate survey they include areas in which they feel that they are lacking in professional development, and where they want to see different things. This year, we used that climate survey when all the administrative team was together, and based off their need we ranked how we wanted to go with the professional developments that we would be using this year.

In regard to using teacher evaluation data to determine professional development, two of the six principals indicated they were taking into consideration the testing data that is now part of teacher evaluation. Specifically, one principal said they were focusing on reading because their scores indicated students were falling behind. Another principal indicated that they do sessions on PVAAS data every year, “We always do a session on PVAAS, at least one.”

One principal was asked specifically, in a follow-up question, about the use of teacher evaluation data for determining professional development. His answer indicated this is not

something they do, “We really didn't do that last year. That wasn't something that we really considered. We really just looked at what their feedback was saying that they wanted, more than anything.”

It was difficult to ascertain how embedded the professional development is or the duration. In retrospect, it would be have been valuable to have an interview question that specifically addressed this. Despite the lack of specificity in the question, some of the principals did allude to this in their answers, and their responses varied. One principal mentioned having “a lot of sessions” on student data. This suggests that the professional development was on-going and embedded since the data was from their students. Another principal mentioned that his teachers collaborate in professional learning communities on areas they have identified. He stated that they have moved away from designated days. This aligns with effective professional development practices. One principal indicated that their professional development is not embedded, but rather is offered on days as determined by the teacher contract.

Overall, the data suggests that a majority of the principals take into account teacher feedback when determining professional development priorities. This aligns with the literature as being an effective practice. The duration and frequency of the professional development varied, but additional data would be helpful in gaining a better picture of current practices.

When looking at how teacher evaluation data is used to determine professional development priorities, several principals alluded to using student data in making those decisions. None of the principals mentioned specifically using feedback from four domains of effective instructional practice when determining priority areas, but it is possible teachers are taking into consideration those areas when identifying their needs. More data would need to be collected in this area to gain a better understanding.

## **5.0 DISCUSSION AND CONCLUSIONS**

### **5.1 PURPOSE OF THE STUDY**

This mixed method study was a practitioner inquiry into the teacher evaluation system that was enacted as a result of Pennsylvania House Bill 1901 (Act 82 of 2012). While Act 82 covered a variety of areas, this study focused specifically on the new teacher evaluation system that was enacted. The study was an effort to gain an in-depth understanding of (a) factors that led to the passage of the Act 82 teacher evaluation policy, (b) what the policy was intended to do, and (c) how the policy has been translated into actual practice. This area of study was deemed important for two reasons. First, research that suggests that teachers have the greatest impact on a child's achievement in school when compared with other factors in the school environment (National Commission on Teaching & America's Future, 1996). Second, resources in the areas of funding, human resources, and technology are required to implement the evaluation system as required by Act 82. This represents an investment of time and money by districts. Therefore, this study sought to understand how the policy has been translated into practice.

#### **5.1.1 Methods and procedures**

This study was designed to be a policy analysis of the Act 82 teacher evaluation provisions through a retrospective review of policy development as well as a prospective analysis of school

principal perceptions and experiences. The practitioner inquiry approach was deemed appropriate because I have worked in the field and have extensive background experience with the classroom observation tool being used in the current model.

This mixed methods study commenced with a retrospective look at factors that led to the passage of the current evaluation model. In the review of literature, the Multiple Streams theory (Kingdon, 2011) was used to analyze historical events and documents leading up to the passage of Act 82. Once the historical look at Act 82 was completed and intended purposes of the policy identified, how the policy was translated into practice was investigated. This was done using a framework focusing on the relationship that exists when a policy moves into practice (Cohen et al., 2007). This relationship can be one of cooperation or conflict. Cohen et al. (2007) identified four factors that tend to determine the outcome of this relationship. Those factors are the nature of the policy, instruments, capability of the practitioners, and the environment. The nature of the policy refers to the level of clarity or ambiguity. The more clarity a policy has, the easier the implementation, but it may also be more restrictive in nature. The more ambiguous it is, the harder it can be to implement as intended. Yet, the ambiguity may provide flexibility allowing the policy to be adapted to the local context. Instruments refer to incentives, ideas, funding, and regulations. Capability refers to how much the policy requires a departure from current practices. The greater the departure, the greater the likelihood of increased incapability of the practitioners to implement the agenda as they grapple with new and possibly unfamiliar practices. The final one is environment, which explores factors in the environment that either support or inhibit the policy's implementation into practice. Using the following research questions, principals' experiences with implementation of the policy was explored.

RQ1: What are school principals' understandings of the intended purposes of Act 82?

RQ2: How have principals been prepared to implement the Act 82 Educator Effectiveness policy in their schools?

RQ3: How have principals implemented the Act 82 Educator Effectiveness policy in their buildings?

The first round of data collection was completed through an electronic survey that was sent to public K-12 principals serving in Beaver or Washington counties. The second round of data collection was done through semi-structured interviews.

### **5.1.2 Limitations**

There were several limitations to the study which must be noted. The first limitation is my own perceptions as the researcher completing this study while working in the field. I currently serve as Superintendent of Schools in a public K-12 school setting located in Allegheny County. At the start of this study, I was serving as the Assistant Superintendent in a public K-12 school setting also located in Allegheny County. As part of my professional role as Assistant Superintendent, I was in the process of planning our district's next phase of implementation of the evaluation model. This involvement led to my own perceptions regarding the positive and negative aspects of the policy, which were addressed through the coding process and subsequent analysis as described in Chapter 3.

The second limitation is low response rate. As was described in Chapter 4, there were challenges with getting principals to respond to the survey. In the end, 28 principals responded with only 27 completing the full survey. This totaled a response rate of 26% of all of the principals invited to participate. For the interview portion of the study, six principals participated, which is approximately 6% of all of the principals invited to participate. While the

data provided insights into the experiences of principals implementing Act 82, the results can in no way be generalized due to the low participation rate. Despite the inability to generalize the findings, the results did reveal areas for further study which are discussed later in this chapter.

## **5.2 MAJOR FINDINGS AND DISCUSSION**

The following section provides a summary of the findings from the retrospective look at what factors led to Act 82. This is followed by findings and recommendations in each of the research question areas.

### **5.2.1 Historical background**

The initial focus of this study was a retrospective look at what factors led to the implementation of Act 82. In the review of literature, a section was devoted to historical events that have had a major impact on education policy. These events were analyzed using the Multiple Streams framework (Kingdon, 2010). This retrospective look began with the report “A Nation At Risk” (ANAR) which was published in 1983. This report brought forth the theme of economic competitiveness on an international level. The report began by describing a rising level of mediocrity in the schools that US citizens needed to be concerned about due to the increase in international competition and the higher level of skills needed by American workers so the country could remain economically competitive. During this time period, media coverage of education policy grew significantly (Dominick, 1984). There was also an increased focus on education policy at both the state and federal levels (Hunt & Staton, 1996).

Several years after the release of ANAR the focus went directly to teachers. The Task Force on Teaching as a Profession was created to examine the teaching profession. A report developed by the task force, *A Nation Prepared: Teachers for the 21st Century* laid out two priority areas. The first was the need for more rigorous academic standards and the second focused on teachers. The report described a need to create “a profession equal to the task-a profession of well-educated teachers prepared to assume new powers and responsibilities to redesign schools for the future” (Task Force on Teaching as a Profession, 1986, p. 2).

In 1994, Congress passed Public Law 103-227 more commonly known as Goals 2000: Educate America Act. This legislation signaled growing accountability as it focused on academic standards and standardized assessments. It also continued the trend of increased federal involvement in education policy. The ANAR report made education a federal level issue by framing it as a national security concern. Goals 2000 indicated that the problem would be addressed at the state level but with federal guidance (Schwartz & Robinson, 2000).

In 2001, Public Law 107-110 titled, *To Close the Achievement Gap with Accountability, Flexibility, and Choice, So No Child is Left Behind* went into effect. This law was an amended version of the Elementary and Secondary Act of 1965 and became more commonly known as No Child Left Behind (NCLB). Accountability was one of the central themes of the law (Hess, 2003; Linn et al., 2002; Peterson & West, 2003; Rudalevige, 2003) and it continued the trend of increased federal involvement. Prior to the passage of NCLB, states had been developing their own content standards leading to inconsistencies (Hamilton et al., 2007; Linn et al., 2002). The new law required all states to demonstrate that their students were making increases in achieving proficiency on these standards. With the passage of NCLB, there was a significant shift in



regards to the federal government influencing state level policy and the theme of accountability continued to grow.

In 2007, a dire forecast was issued for the US. According to researchers at the Educational Testing Service (ETS), the country was in the midst of a perfect storm with a grim outlook (“America’s Perfect Storm,” para. 1). A report issued by ETS entitled *America’s Perfect Storm: Three Forces Changing Our Nation’s Future* (Braun et al., 2007) outlined three converging forces that put the country on a perilous course towards disaster: divergent skill distribution, a changing economy due to globalization and changing demographics within the US. These three forces interacting was considered cause for great concern because the prediction was that over the next 25 years more highly educated Americans would be leaving the workforce at a time when jobs increasingly required high level skills. According to the report, those entering the workforce would do so less educated and therefore less prepared to meet the demands of the changing economy. This placed the nation in perilous risk considering the rising competitiveness of other countries. This report led to an increased national focus on developing sound educational policies.

During the 2000s, the three streams identified in Kingdoms Multiple Stream frameworks came together opening a policy window for Race to the Top which eventually led to the development of the *Educator Effectiveness* policy enacted through Act 82 (Pennsylvania Department of Education, 2014). The problem stream described in the previous paragraph predicted a dire future for the US due to the lack of growth of a highly qualified workforce. In the policy stream, the Brookings Institution published a report in 2006 titled *Identifying Effective Teachers Using Performance on the Job* authored by Robert Gordon, Thomas Kane, and

Douglas Staiger (2006). In the political stream, a new President was elected in November of 2008.

The report from the Brookings Institute focused on increasing teacher quality through five recommended actions: create non-traditional pathways for people to enter the teaching profession, make it more difficult for teachers to earn tenure including linking student data to specific teachers, provide financial incentives to highly effective teachers willing to work in areas with a high proportion of low income students, use of multiple measures in teacher evaluation and development of data systems to link student performance with individual teachers in order to measure effectiveness. These recommendations were later incorporated into Race to the Top, a competitive grant process used by the executive office in an attempt to circumnavigate the issues occurring with the reauthorization of NCLB due to Congressional gridlock (McGuinn, 2006). Pennsylvania applied for and received funding through this grant program. These funds were targeted towards school reforms designed to improve student achievement. The Pennsylvania Department of Education (PDE) implemented the *Educator Effectiveness Project* (Pennsylvania Department of Education, 2014) that included a teacher evaluation model based on multiple measures. The stated goal for the Educator Effectiveness Project was, “To develop educator effectiveness models that will reform the way we evaluate school professionals as well as the critical components of training and professional growth.” (<http://www.education.pa.gov>)

## **5.2.2 RQ1: Understanding principal perceptions of intended purposes**

The first research question, “What are school principals’ understandings of the intended purposes of Act 82?” was designed to compare principals’ understandings of intended purposes to what was discovered in the review literature as being the actual reasons for the policy. In order to

answer this question, a review of the literature was completed to determine the intended purposes. The first section provides background on the intended purposes of Act 82 *Educator Effectiveness* as identified through the review of literature. The second section discusses principals' perceptions of the intended purposes and how they align with actual intended purposes.

#### **5.2.2.1 Intended purposes identified in the literature**

Beginning in the 1980s, the theme of teacher accountability and quality increased in importance. During this time period, concerns grew about the workforce in this country and the ability to remain competitive in an increasingly globalized world. Several trends emerged: increasing federal involvement in education policy and accountability. The second trend was shown through the increased focus on rigorous academic standards, the use of academic data to make decisions and eventually the use of student data in teacher evaluation.

There were five intended purposes of the policy that were identified in the literature. All of them fall under the realm of "accountability" except for the final one. In general, accountability refers to the focus on how schools perform and enacting rewards and/or consequences based on that performance. An article published in *Education Week* in 2004 described the increasing focus on school accountability, "Accountability—the idea of holding schools, districts, educators, and students responsible for results—has become the most-recent watchword in education. The push for accountability has grown out of a common perception that states traditionally monitored the "inputs" in public education—such as the number of books in the school library or the number of computers in the classroom—but paid too little attention to performance. In the 1980s, the nation's governors proposed a kind of "horse trade": The state

would provide more flexibility in how schools operated, as well as more money for schools, if educators would agree to be held more accountable for student achievement.”

The first intended purpose of Act 82 was ensuring high quality teachers which I included under the realm of school accountability, because the teacher has the greatest impact on student achievement as was described in Chapter 2. The focus on high quality teachers can be traced back to the report, *A Nation Prepared: Teachers for the 21st Century* with a focus on attracting high quality candidates to the field by increasing salaries and career opportunities. Subsequently, the focus on high quality teachers was described in the 2006 report *Identifying Effective Teachers Using Performance on the Job* published by the Brookings Institute. In 2009, President Obama gave a speech outlining the main tenets of Race to the Top. One of those main tenets was placing “outstanding teachers” in the front of everyone classroom. The Act 82 *Educator Effectiveness* project resulted from Pennsylvania’s receipt of Race to the Top funding and was therefore focused on ensuring high quality teachers.

The second intended purpose was the ability to remove ineffective teachers. This also falls under accountability since it was meant to be a consequence of not producing positive student achievement results. The Brookings report specifically recommended making it harder for teachers to earn tenure. Traditionally, teachers would earn tenure after three years of full-time teaching experience. Student performance was not necessarily considered when determine tenure. The Brookings report recommended student data be used in making this determination with the idea that underperforming teacher could be evaluated out of the profession.

The third and fourth intended purposes are related. The third was an evaluation system using multiple measures, and the fourth was using student achievement in evaluation. This is a clear accountability measure since it ties student achievement to how a teacher is rated in the

evaluation process. Both of these recommendations came directly from the Brookings report, but the increased focus on looking at student achievement data can be traced back to No Child Left Behind.

The final intended purpose is a focus on professional development and training. This is specifically mentioned in the goal for *Educator Effectiveness* which states, “To develop educator effectiveness models that will reform the way we evaluate school professionals as well as the critical components of training and professional growth” ([www.education.pde.gov](http://www.education.pde.gov)). For the purposes of this study, I did not include this under the category of “accountability”, but that does not mean that one could not make a case for including it.

#### **5.2.2.2 Alignment of principal perceptions**

The following two sections describe principals’ perceptions of the intended purposes, where they align and where they diverge. The third section talks about how principals have taken a tool that was designed as an accountability measure and have used it help promote teacher growth.

#### **5.2.2.3 Accountability**

As was stated in Chapter 4, the data suggests that the participants’ understandings of the intended purposes generally align with the intended purposes. Specifically, they see this policy resulting from increasing accountability in education which the literature suggests is correct. Three themes that emerged were: accountability, effective evaluation tool and consistency. The predominant theme that emerged was accountability which aligns with intended purpose. There were two sub-themes within accountability: focus on student achievement and identifying ineffective teachers. Both of these directly align with the intended purposes as identified in the literature. What was interesting is only 16% of the 19 principals who responded to this question,

identified these specific accountability measures (student achievement and identifying ineffective teachers). While it is impossible to definitively conclude why this occurred, it is possible that principals see the policy as an accountability measure but have not given it further thought.

I noticed during the interviews, when I asked the question about intended purposes the principals typically needed a minute to think about this. It seemed to suggest that this was something they had not previously given a lot of thought to. They seemed to understand the policy was designed for accountability reasons, but they had not thought much beyond this. What was interesting was they did seem to give a lot of thought to how to make the tool beneficial for their context. This was suggested by the data in the area effective evaluation tool, and by their responses to RQ3 which are discussed later in this section.

#### **5.2.2.4 Effective evaluation tool**

The theme of an effective evaluation tool suggested somewhat of a divergence from the intended purpose of the policy. This was especially noticeable considering their responses in RQ3. Their responses showed alignment with the intended purpose of ensuring a high quality teacher, but their answers indicated this would happen by using the evaluation tool to support teacher growth. The literature suggests that the intent of the policy was to ensure high quality teachers by evaluating out underperforming teachers. The principal answers did not reference this. In fact, the new tool has made it more difficult to rate a teacher as needs improvement, because of how the mathematical calculations work out. It is possible principals are aware of this and therefore have chosen to focus on using the tool for teacher growth, but that cannot be concluded from this study. It is interesting that the principals focused in their answers on using the tool for teacher growth, but only one principal specifically referenced professional development and training for

teacher growth. This is ironic since the goal of *Educator Effectiveness* specifically references professional development and training.

#### **5.2.2.5 Teacher professional development**

The historical review and intended purposes of Act 82 show an increasing focus on educational accountability. The responses provided by the principals indicate that they have a general understanding of Act 82 being an accountability measure. Not emphasized in the historical review of literature is a focus on professional development and training for teachers. While teachers are expected to perform at high levels, there has not been an emphasis on providing the support to do this. The goal of Act 82 states that professional growth and training are a focus, but this study did not find evidence of this. What was discovered was principals taking a tool designed for accountability purposes and using it to help promote teacher growth. As was stated in Chapter 2, if a person is not in agreement with a policy's aims, the challenge of implementation will be that much greater. Interests refers to whether or not the practitioner has a vested interest in the policy implementation (Cohen et al., 2007). The results of this study suggests that principals seem to have more of a focus on teacher growth as opposed to accountability as described in the literature. This is not meant to suggest the principals did not agree with holding teachers accountable. On the contrary, their responses suggested some level of agreement with this. Where they diverged was they seemed less focused on consequences and more focused on helping teachers grow professionally. A response to the survey question capture this well:

The Danielson model was not designed to evaluate teachers. It was a tool designed to help teachers focus on specific areas to grow. The state's model places such high stakes

on goal setting that teachers choose goals that are easily achievable. This was not the purpose of the Danielson's tool.

Overall, of the principals who provided a response to this questions and/or participated in the interview, their responses seem to indicate that they understand Act 82 as a policy born out of an increased focus on accountability. While their understanding aligns with intended purposes, those purposes do not appear to meet their needs as they have shifted the focus of the tool during implementation. Instead of using it as an accountability measure, they have utilized the four domains of effective instruction and the student data to encourage professional growth in their teachers. It must be kept in mind that some of the principals did not respond to the open-ended questions, and therefore may have had a different experience than what is captured in these results.

### **5.2.3 RQ2: Understanding how principals were prepared**

This part of the study sought to understand how principals were prepared for the implementation of Act 82. This was looked at through the framework of policy instruments, specifically training, and capacity of the principals. It appears that there was an early focus on providing training for principals on the new evaluation tool. The majority of the principals who responded indicated that they participated in IU and district trainings. They also indicated that these trainings were effective in helping them feel prepared for the implementation. A smaller group of principals indicated that they participated in some other form of training (i.e. PLCs, web-based, independent learning) which were also effective in preparing them. What was interesting to find was only 1-2 principals indicated the trainings they participated in were “on-going” with the exception of the category independent learning activities. In this area, eight principals indicated



it was on-going. Considering what has been described in the literature about adult learning theory and best practices in professional development, this is an area to consider further and is discussed more in-depth in following sections.

In the area of capability, the principals indicated that time working with the evaluation tool helped build their capability which is not surprising. It was less clear how helpful prior experience with a similar tool was in helping principals feel prepared for implementation. In the survey, 11 principals indicated that they had worked with a similar tool and out of this group 43% indicated that this prior experience helped them with implementation. In the interviews, prior experience was a prevalent theme, but when interpreting the results it should keep in mind that the interview participants may be included in the survey results. The principals who were interviewed indicated that they felt their prior experience had a positive impact on helping them feel prepared for implementation. It was also suggested that their colleagues, who did not have prior experience, found implementation more challenging. Part of the discrepancy may have to do with the kinds of prior experiences principals had with the tool. It is possible that the ones who indicated it was not effective had a different experience than the ones who indicated it was effective. This study did not provide an in-depth probe into prior experience. Due to the low number of principals responding in the survey, it is impossible to make conclusions, but it is worth discussion and future study.

#### **5.2.3.1 Intersection of capability and instruments**

These results suggest a combination of instruments and capacity building may have helped principals implement the current evaluation tool, but they also suggests that a focus on training has not continued in most districts. Considering what the literature says about adult learning and professional growth, it seems that continuing to provide training as principals gain more

capability with the evaluation tool could be very beneficial, especially if the training moves beyond being introductory in nature. The data indicated that the trainings initially provided by the IUs were helpful, especially to those who felt “somewhat prepared” versus those who felt prepared.

As a practitioner, I attended these trainings in one of the IUs that was covered in this study. The trainings tended to be more of an overview of the four domains as opposed to going in-depth which was appropriate for initial implementation. The data also suggested that principals gained more capability as they worked with the tool. It has now been 5 years since the tool was implemented giving principals more time to develop their capability. It seems that it would be advantageous to continue to have more sophisticated trainings as practitioner capability grows, but this has not occurred in the districts covered in this study. The principals reported that trainings have not been on-going with the exception of some principals engaging in on-going independent learning activities. The next section talks about adult learning theory and effective professional development as they relate to Act 82 implementation and on-going support of the main implementers, the principals.

### **5.2.3.2 Adult learning theory**

In this section, an overview of adult learning theory is provided. The main focus areas of the theory are then applied to principal experiences with Act 82 implementation. During 1960s, Malcom Knowles brought attention to the field of adult learning as he began to synthesize the concept of andragogy, adult learning theory, making it more widely known through his writings (Taylor & Kroth, 2009). The andragogical model developed by Knowles and associates focused on the following assumptions: adults need to know why they need to learn something before attempting to learn it, adults’ self-concept is based on being independent so self-directed learning

should be a goal, the learner's life experiences should be taken into account and emphasized, what is being learned should be useful to the learner in coping with current or soon to exist life situations and adult learners respond more to internal motivators as opposed to external rewards (Knowles, Holton, & Swanson, 2012).

The first part of the model is knowing why an adult needs to learn something. Within the context of Act 82 implementation, the principals already know why they need to learn more about the evaluation tool as it is required by law. What is unknown is if they are motivated to learn more. Further research could address how they value the tool to determine if they have the motivation to learn how to use the tool more effectively or if they see it simply as a policy with which to comply. The survey answers suggested both. Some of the responses indicated that principals saw it as a tool for promoting teacher growth while other answers seemed to focus more on the amount of required paperwork.

The next two aspects of the model are self-directed learning and life experiences being taken into account. This is interesting since self-directed learning was the only type of learning that was found to be on-going and the responses indicated that the choices were being driven by the principals themselves. Moving forward, districts would be wise to provide on-going professional development opportunities knowing that they may vary from district to district since contexts vary. The principals should also be active participants in providing input into the content and nature of those opportunities since they are the ones most knowledgeable about their experiences with the evaluation tool. This is also supported by the literature on best professional development practices which is covered in the next section.

Finally, the learning should be useful to the learner in coping with a current life situation. This applies to all of the principals since the evaluation tool is a required part of their jobs and

relates to a potential internal motivator. The principal interviews indicated a focus on using the tool to promote teacher growth and not a focus on it as an accountability tool. It may be that the principals feel more of an internal motivation to use the tool effectively to help their teachers succeed. This motivation, combined with professional development opportunities that help them succeed in using the tool effectively, could potentially have a positive impact on the effectiveness of their teachers resulting in greater student achievement. In the end, that is one of the goals of Educator Effectiveness, but there needs to be a continual focus on the use of the tool beyond initial implementation training.

#### **5.2.3.3 Characteristics of effective professional development**

The data from this study suggests that there was an early focus on training and preparing principals for implementation of Act 82, but that focus has not continued. As principals have gained more experience with the evaluation tool, they have increased in their capability as compared to when they began. If districts were to return to the focus on policy instruments, specifically high quality professional growth opportunities, they could potentially increase the effectiveness of both their principals and teachers. Considering what the literature says about the impact of a teacher, this could potentially have a positive impact on student achievement. In the literature, there are a number of practices that are considered to be effective when delivering professional development that is intended to change practice and impact student achievement. The first consideration is duration. Professional development (PD) should extend over time allowing educators to go in depth with the topic and it needs to include plans for ongoing follow up (Garet et al., 2001; Porter et al., 2000). The traditional method of the one day conference presented by an expert outside of the classroom has been widely criticized in the literature

resulting in an increase in more “reform” style PD activities such as study groups, mentoring and coaching (Garet et al., 2001).

Another component of professional development is educators having some choice in the topic (Porter et al., 2000) so it will be specific to their area of need. For example, in one of the interviews a principal referenced that it would be helpful to have additional training on evaluating non-teaching professionals. This concept extends to the idea of PD being job embedded (Garet et al., 2001; Hiebert, 1999; Porter et al, 2000) which directly aligns with the theory of andragogy focusing on learning being something that the adult learners can apply to their everyday situations in the classroom. Job-embedded learning is the practice of educators sharing information, formally or informally, about what works and does not work, trying new practices and reflecting together (Zepeda, 2012).

The critical piece here is educators having opportunities for active collaboration (Garet et al., 2001; Porter et al., 2000). This type of PD provides educators an opportunity to engage in the transformational learning process by discussing and working through common issues. In the interviews, the majority of the principals referenced professional conversations as being valuable in their understanding and implementation of the evaluation tool. In some cases, these conversations occurred as part of a planned training, but more often they happened during informal conversations. Providing on-going opportunities for principals to gain new knowledge as they gain in capacity and have professional conversations about their work could be a powerful tool for districts. Linda Darling-Hammond (1995) sums it up well when she states that “Like their students must do, teachers also construct their own understandings by doing: by collaborating, by inquiring into problems, trying and testing ideas, evaluating and reflecting on the outcome of their work” (p. 24). She goes on to say that by providing educators with

opportunities to work on real problems of practice, educators are able to develop a deeper understanding of their practice as well as to develop an increased sense of ownership.

For districts interested in progressing their professional learning to the next level, the *Standards for Professional Learning* (Learning Forward, 2013) provide a guide with the necessary components. These standards were specifically titled “professional learning” as opposed to “professional development” to put an emphasis on the active role the educator plays in this process. The standards are divided into seven categories: learning communities, leadership, resources, data, learning designs, implementation and outcomes. The first standard addresses the need for educators to work in learning communities that meet continuously, share collective responsibility and maintain a focus on their goal(s). The second standard covers the allocation of resources which includes costs of human resources, time for learning, technology and material resources. This also includes the coordination and fair allocation of resources.

The next standard is data. In order to know where the needs are in a school, data needs to be collected, analyzed, used in planning and routinely evaluated. Data sources also need to be rich and varied. The fifth standard covers learning designs. This standard emphasizes the need to incorporate the theories of adult learning when planning for professional learning. The next standard is implementation. Quality professional learning requires sustained effort and this standard talks about the need to plan for on-going support. The final standard looks at outcomes that are based on high expectations. In this standard, professional learning is directly tied to increased student outcomes.

#### **5.2.3.4 Significance of continual professional learning for principals**

In Chapter 1, the rationale for why this study was important was stated. The first reason that the research suggests is teachers have the greatest impact on a child’s achievement in school when

compared with other factors in the school environment (National Commission on Teaching & America's Future, 1996). The second reason is the amount of resources in the areas of funding, human resources, and technology required to implement the evaluation system as required by Act 82. In addition, early feedback indicated that the evaluation process is time consuming when completed as intended. Considering the investment of resources and the high stakes involved for children, it is important the focus on how this policy is being implemented continue.

A lot is on the line with the implementation and continued influence of this policy in schools. At the start of implementation, there was a focus on preparing the primary implementers. Yet, as time has passed there does not appear to have been a continued focus on ensuring effective use of the evaluation tool that was enacted because of this policy. Considering the amount of time and money invested in this policy, it would be wise for districts to focus on increasing the effectiveness of the evaluation tool by continuously developing the capabilities of the principals. Now that principals have developed some capability with the tool, districts have a prime opportunity to provide professional growth opportunities that could potentially have a positive impact on helping principals use the tool to promote educator effectiveness. The information in the previous sections provides the knowledge to do so in a manner that is grounded in research.

#### **5.2.4 RQ3: Implementation of the policy**

RQ3 was designed to learn about how Act 82 has been implemented in the schools. In the policy literature, this falls into the category of environment. As was discussed in Chapter 2, an organization can be structured in such a manner that capability to implement a policy is either enhanced or constrained (Cohen et al., 2007):

Policies that cause practitioners to change practice, especially in a major way, creates a certain level of incompetence. Organizations can impact how that incompetence is addressed as they enhance capability by enabling and institutionalizing the exchange of knowledge about practice and the interests, values, and knowledge that inform practice. (p. 540)

In this study, principals were inconsistent in their perceptions of whether or not this policy had the intended effect of ensuring an effective teacher in every classroom. This indicates that there is still work to be done in meeting this stated goal.

The following sections offer suggestions on what districts could do to support this policy in their districts in order to support goal of ensuring effective teachers. The following sections addresses further the continued professional development of principals and recommendations on focus areas.

#### **5.2.4.1 Addressing principal capability**

The first area of RQ3 focused on effective evaluation practices. When describing their work with this tool, principal's responses ranged from negative to positive. The negative comments reflected a tool that is time-consuming, paperwork intensive and not effective. Positive comments revolved around the tool providing opportunities for professional dialogue collaboration and reflection. Those who were in a position where this signified a drastic change of practice may be the ones responding the most negatively. The interview responses support this in that this group of principal generally described the implementation in positive terms, and based on what they shared about their prior experiences, this policy was not a significant departure from prior practice. The theme that was most prevalent throughout all of the



interviews was the idea that the new model has increased professional dialogue, collaboration and reflection.

Every participant touched on this idea. The participants reported more discussions focused on instruction with their teachers as well as teachers dialoguing with each other about effective practices. While the data suggests initial competency may be impacting principal perceptions and implementations, it cannot be absolutely concluded this is the case without further research into the experiences of the principals who describe the tool in negative terms. As was previously mentioned, the principals who participated in the interviews were generally positive in their responses suggesting that principals who had more negative experiences may have self-selected out of the interview portion of the study.

Regardless of being inconclusive, districts could potentially benefit from providing professional growth experiences for principals to help them develop their competency with the tool. In the section on RQ2, the need for on-going opportunities was discussed and this section expands further on that topic. In the area of evaluation practices, I looked for examples of effective practices i.e. effective tool, trained observers, multiple observers for each teacher and use of student data. Considering these components, RQ3 provided insights into areas that may prove valuable in regards to areas to focus professional training. The following recommendations fall under the areas of effective tool, trained observers and use of student data.

The survey and interview data indicated that all of the principals went through some type of initial training. Yet, when describing the evaluation tool, there were mixed feelings about its effectiveness and value. The interviews proved interesting in that those principals were generally positive about the tool and seemed to feel comfortable using it. This group of principals offered insights into what they found to be valuable about the tool, and they also

offered insights into areas they would like to continue to grow. Therefore, this section of recommendations comes mainly from the data that was collected from the interview participants.

#### **5.2.4.2 Professional growth opportunities**

The first recommendation, which was covered in RQ2, is providing on-going professional growth opportunities. As was stated previously, there was an initial push to train principals, but that focus has not continued. The literature talks about the challenges of implementing a policy that is a departure from current practice. For the interviewees, their comments suggested that their prior experiences helped them have a successful implementation, but they indicated that some of their colleagues struggled. While it is probable that all of these principals now understand how to use the tool, the effectiveness of it is impacted by the principal's competency. Continuing to develop all principals' capability would likely continue to increase the effectiveness of the tool. Even within the group of principals interviewed, only four said they thought the evaluation tool had an impact on instruction, one said it was a start and the other said "not necessarily" suggesting even those principals experiencing some success still need professional growth opportunities. The principals' comments in the interviews also provided guidance in where districts may want to target those professional opportunities.

#### **5.2.4.3 Professional dialogue**

The theme that was most prevalent throughout all of the interviews was the idea that the evaluation tool has increased professional dialogue, collaboration and reflection. Every participant touched on this idea. The participants reported more discussions focused on instruction with their teachers as well as teachers dialoguing with each other about effective practices. Districts could potentially benefit by offering on-going professional development

focused on the four domains and how to have effective professional conversations around them. In the literature, Danielson (2016) reflected about how powerful this can be:

Of all the approaches available to educators to promote teacher learning, the most powerful (and embedded in virtually all others) is that of professional conversation. Reflective conversations about practice require teachers to understand and analyze events in the classroom. In these conversations, teachers must consider the instructional decisions they have made and examine student learning in light of those decisions. (p. 5)

She goes on to describe how, historically, teachers have been very passive when participating in a post-observation conference with the principal. The conference was an event to be endured. In this context, the evaluation tool may be high quality, but when used in this manner it is unlikely to change practice. Having principals who are skilled in the art of professional conversations, using the same tool, is more likely to have an impact.

#### **5.2.4.4 Teacher reflection**

In the interviews, the principals also touched on the idea of teacher reflection. They indicated inconsistent responses from their teachers. For some teachers, their answers to questions within the domains indicated a lot of reflection on practice. Others indicated very little and in some instances teacher comments indicated they only wanted to do what was needed to achieve proficiency. While it is possible some teacher's attitudes will not change despite principal efforts, the effective use of professional conversations can promote teacher reflection even beyond the discussions. Danielson describes this process as "the value of professional conversations extends far beyond the particular settings in which they occur; that is, they have value both in the moment and over time. By participating in thoughtful conversations about practice, teachers acquire valuable habits of mind that enable them to pursue such thinking on

their own, without the scaffolding provided by the particular conversation. On another occasion, teachers can consider the lessons they have extracted from a given situation and determine their applicability to a new set of circumstances. It is this transfer of insight that makes professional conversation such a powerful vehicle for learning” (Danielson, 2016, p. 5).

#### **5.2.4.5 Discussions about data**

Another area districts can focus on is helping principals develop their capability in using data to inform instruction. This includes how to talk with teachers about data, understanding which data is useful and how to help teachers translate data into information that informs instruction. In his book, *How Teachers Can Turn Data Into Action*, Daniel Venables talks about how teachers and administrators often find discussions about data unpleasant. In this study, that observation showed up in the interviews. The principals reported teachers having mixed feelings about the student data piece, but two themes emerged from their answers: inequity and control. Inequity referred to the teachers’ perception that some teachers are held more accountable under this system based on what grade level and subject area they teach.

The idea of control refers to the level of control teachers perceive having on the outcome of their data. The higher level of control they felt they had the less negative their feelings were about having data included. While there are aspects of this evaluation tool that teachers and principals have no control over, such as how the data is used to calculate ratings, they can employ practices to use the data effectively to inform instruction.

During the interviews, principals were asked how they talk with teachers about data. Their answers varied which again is not a surprising outcome. There was also inconsistency in which data principals focused on. As was mentioned previously, I attended the IU trainings that were offered when this policy was being implemented. During those trainings, the data piece

was not covered beyond explaining how the calculations were completed for the final ratings. Considering some of the negative terms that were used to describe the teachers' feelings about data, it would be beneficial for principals to have the skills to talk about data and understand how to put in structures in place that effectively engage teachers in the use of data which teachers perceive as being beneficial to their work. This could potentially help teachers feel a sense of control and build in them the competencies to utilize data effectively. In the end, teacher engagement in the use of data is much more important to school success, and the effectiveness of this policy, than compliance with mandates (Coffman, Gonzalez Molina, & Clifton, 2002; Reeves, 2004).

#### **5.2.4.6 Multiple observers**

The final evaluation practice I looked at was multiple observers. The literature suggests using multiple observations and observers for each teacher (Darling-Hammond, 2013; Gates Foundation, 2012). The data from this study suggests that this is not a common practice which is consistent with my own experience as a practitioner in the field for over twenty years. If multiple observers are used, it is typically in response to a teacher who is struggling. Having different administrators observe each teacher is a practice that districts can implement without additional financial costs. Considering the value in professional dialogue this may provide a benefit to teachers, especially if their principal is less adept at having these conversation. This practice could also be used to build capability in principals who are not skilled in this area. By pairing an inexperienced or struggling principal with a stronger colleague, that principal has a professional growth opportunity embedded into the work day. They can observe together and have follow-up discussion about what was observed and how to most effectively communicate with the teacher.

#### **5.2.4.7 Software**

One thing that came out of the interviews was the use of software to manage the evaluation process. The principals described this as being helpful. Considering some of the negative feedback in the survey regarding the amount of paperwork and logistics involved in this process, it may be beneficial for districts to invest in this if they have not already done so. This is also a possible area of further research since this is not an area this study covered. The use of software, as a policy instrument, could be researched to determine if it has an impact of how effectively the policy is implemented in districts.

### **5.3 IMPLICATIONS AND RECOMMENDATIONS**

The results of this study suggest that the current teacher evaluation tool has the potential to positively impact the effectiveness of teachers. A key factor in this is the capabilities of the principals. The data suggests that when the policy was initiated there was a focus on providing the supports needed by principals for a successful implementation, but this focus has not continued which may potentially undermine the effectiveness of this evaluation tool. This is especially concerning considering the impact a teacher has on a child's success in school. In order for this policy to work as intended, there needs to be a commitment by legislators, the Pennsylvania Department of Education and districts to continue to provide the supports needed. Otherwise, Educator Effectiveness will be yet another initiative that goes by the wayside. In the following sections, I offer my own perspective on the challenges that have occurred in implementing the policy and recommendations on how support can be provided at both the state

and district levels. These recommendations are based on my professional experiences as well as what is detailed in the literature.

### **5.3.1 Implementation challenges**

Implementing a new teacher evaluation tool is a monumental task. As this study revealed, principals who had greater capability were able to put into place professional practices that have the power to enhance teacher effectiveness: professional dialogue, collaboration and reflection. Other principals who had less capability found the implementation more challenging. The data suggest that the issues of capability were addressed early on with trainings at the IUs and districts. Unfortunately, this effort did not continue. While this study did not seek to understand why this did not continue, my own experience suggests that it may have resulted due to competing factors, especially the concurrent implementation of the Pennsylvania Core Standards.

For districts and principals, the evaluation tool enacted as a result of Act 82 represented a major policy implementation. This study focused mainly on the 50% classroom observation portion, but also included in the evaluation was student data. This was the first time student data was specifically tied to teacher evaluation. This alone represented a significant shift in practice. Unfortunately, the state also rolled out another significant change in practice which was the implementation of the PA Core Standards. This change caused a dramatic decrease in student scores at approximately the same time student scores were being included in teacher evaluation for the first time.

In 2010-2011, the state began sending out information on the new standards. In 2011-2012, trainings began on the new standards. This coincided with the enactment of Act 82. In 2012-2013, districts were expected to begin aligning their curriculum with the standards and

preparing for implementation. This represented a significant effort in regards to preparation of teachers and principals in understanding the new standards. It also represented a significant change in instructional practices for teachers. For districts, it also included reviewing and investing in updated resources that aligned with the new standards. In 2015, the PSSAs were fully aligned to the PA Core Standards. Throughout the state, there was a noticeable drop in achievement scores as districts worked to recalibrate to the new standards.

Essentially what policymakers had done was create the perfect storm. The literature suggests that when a policy is moving into practice the greater the departure from current practice the greater the conflict as the implementers struggle to enact it. This can be addressed through effective use of policy instruments. Unfortunately, in the case of Act 82, the legislature created a situation where two policies, both representing significant shifts from current practice, were being enacted concurrently. Even more significant was the impact PA Core Standards had on teacher evaluation during the early years of its implementation. For the first time, teachers were being evaluated on student data. This occurred at a time when new standards were being enacted and consequently caused achievement scores to drop significantly. Based on my professional experience and what I discovered in the literature, this created a situation where it would be incredibly difficult to implement both policies effectively.

A second challenge that has occurred is the inability to remove ineffective teachers using the current evaluation tool. This is ironic since one of the intended purposes was to remove ineffective teachers. The Pennsylvania Department of Education provides historical data on how many teachers are rated unsatisfactory each year. During the 2013-2014 school year, the first year of the rating system, 98.2% of all teachers were rated satisfactory. This was the highest percentage in 5 years. What administrators have discovered with the evaluation tool is it is



almost mathematically impossible to rate a teacher unsatisfactory. This is especially true in a high performing school district where the overall achievement scores can help buoy an underperforming teacher. As a practitioner, I heard numerous warnings about this from lawyers as the current tool was being rolled out. They directly stated that it would be virtually impossible to fire a teacher under the current evaluation tool for a variety of reasons, including how the math calculations were configured. This has been a source of frustration for principals as they feel they are stuck with underperforming teachers at a time when accountability is high. This problem came to light early on in the implementation yet there has been no efforts by the legislature to correct it. It is possible that this is why principals have instead focused their efforts on using the tool to help teachers grow professionally.

### **5.3.2 Recommendations**

Based on the findings from this study, what is reviewed from the literature and my own experiences as a practitioner I offer the following 3 recommendations: more consideration from policymakers on policy implementation timelines, continued trainings offered by the IUs, and on-going support within districts.

*Recommendation 1:* The enactment of Act 82 and the resulting teacher evaluation tool created a significant departure of practice for some principals. This is not an insurmountable hurdle, but it is one that requires sustained effort in the form of trainings and supports. Districts in Pennsylvania are limited in both their financial and human resources. Implementing a new evaluation tool would be expected to put a strain on these resources, but it was one that could provide valuable returns as teachers gained in their effectiveness. Unfortunately, the decision was made to layer on another, very significant, policy by concurrently implementing the PA

Core Standards. In the future, policymakers need to consider timelines of major policy changes. Every effort should be made to ensure multiple policies, requiring a departure from current practice, do not go into effect concurrently. Time and resources must be devoted to the implementation to any new policy that is expected to produce significant outcomes.

*Recommendation 2:* The data from this study suggests that policy instruments and capability have an interactive relationship. As principal experience with the tool increases so do their capabilities. This provides an opportunity for IUs to develop more sophisticated trainings beyond the introductory ones that occurred early in implementation. As principals gain familiarity with the tool, they are cognitively more available to gain from advanced trainings. This study suggests trainings that would likely be valuable are as follows: effective use of questioning to promote discussion and reflection, viewing videos of lessons and discussing evidence with other principals, effective conversations around data, evaluating non-teaching professionals, etc. The options are limitless, but, as research suggests, should include input from the principals on what would be most helpful and a focus on collaboration, dialogue and reflection.

Another training issue IUs need to address is new principals. Currently, no trainings are available on the evaluation tool which is concerning for principals who are new in their positions. While they are likely to be introduced to the tool during their preparation programs, assuming they attend school in Pennsylvania, there is currently no IU support once they enter the field. This is a missed opportunity as they begin working with the tool and increasing their capability. The introductory sessions that were previously offered should continue to run at least twice a year to ensure new principals have access.

*Recommendation 3:* The literature on policy to practice suggests that the environment can support or inhibit successful policy implementation. There are actions district can take to provide continued support to principals. A common theme that came out in the data as being beneficial was the opportunity to talk with other administrators. It was also found that most teachers were not observed by more than one administrator. Therefore, it is recommended that districts have principals do observations together. This provides a professional growth opportunity as principals have an opportunity to dialogue about what was observed. It also serves to have multiple people observing teachers which is supported in the literature. This practice should be embedded and on-going. Another suggestion is to encourage independent learning opportunities. The data, while limited in this area, suggested these activities were beneficial and sustained. Districts should support this by providing time for these activities and providing resources such as books, articles, online training, etc. Again, the nature of the activities should be driven by the needs articulated by the principals.

### **5.3.3 Concluding remarks**

The area of policy and policy development tends to be one that educators are not particularly knowledgeable in yet it impacts the work they do on a daily basis. As was described in this study, the involvement of legislators has grown through the years, and it appears this trend will continue. Therefore, it behooves educators to have an understanding of why certain issues rise to prominence while others fade, be informed about potential and existing policies, understand the political environment and most importantly advocate on behalf of the students and communities they serve. Act 82 was implemented through the Pennsylvania legislature. This is significant

because it shows the legislature's willingness to develop policy that has a direct impact on the work educators do on a daily basis.

In the case of Act 82, it is law and so it is in the best interest of children that districts work to ensure it is being done effectively. While educators may disagree on how effective the evaluation tool is, the research and this study suggests that it can bring value to the observation process through the promotion of professional dialogue and conversation. As has been discussed, there are a number of practices that can be employed by districts to help facilitate and strengthen this process. In the future, should the legislature begin work to revise the teacher evaluation process, educators who are well-informed and willing to advocate can have an influence on future policy. It is my hope, that this dissertation has provided educators the motivation and some of the knowledge needed to have a voice in education policy. Our children are depending on it.

## APPENDIX A

### SURVEY LETTER

Dear \_\_\_\_\_,

My name is Caroline Johns and I am a currently a student at the University of Pittsburgh completing a doctorate in Administrative and Policy Studies. I am conducting a research study entitled: *A Mixed Method Exploration of the Efficacy of the Teacher Evaluation System in Western Pennsylvania*. This is a policy study looking at how the intended purposes of *Educator Effectiveness* (Act 82) have been translated into practice. This study has been approved by the University of Pittsburgh Human Research Protection Office pursuant to the U.S. Federal Government Department of Health and Human Services (2009) regulation 45 CFR § 46.101.

I am writing to request your participation in this study by completing a short survey. Your participation in this research project is completely voluntary, and there are no known risks beyond those encountered in everyday life. In the survey, you will be asked about your experience with implementing Act 82 *Educator Effectiveness* as well as your thoughts about its effectiveness as a teacher evaluation tool. All information is kept confidential, and you may choose to exit the survey at any time. The survey takes approximately 10 minutes to complete. To begin the survey, please click on the link below. By clicking on the link below you are acknowledging that your participation in this research is entirely voluntary.

If you would like to receive a summary of the findings at the conclusion of the study, please email your request to [cej19@pitt.edu](mailto:cej19@pitt.edu). Your participation in this research is greatly appreciated!

Sincerely,

Caroline Johns

## APPENDIX B

### INTERVIEW INVITATION TO PARTICIPATE AND CONSENT FORM

Dear \_\_\_\_\_,

I am a student at the University of Pittsburgh working on a doctorate in Administrative and Policy Studies. I am conducting a research study entitled: *A Mixed Method Exploration Of The Efficacy Of The Teacher Evaluation System in Pennsylvania*. This is a policy study looking at how the intended purposes of *Educator Effectiveness* (Act 82) have been translated into practice. This will be done through review of the literature, survey, and interviews of currently practicing K-12 principals in Beaver and Washington Counties.

Your participation will involve a semi-structured interview that will take approximately thirty minutes. A short follow up interview or phone call may be necessary for clarification of your comments if needed. The interview will be audio recorded for accuracy. Your participation will be kept confidential, your name will not appear on any document related to the study, and your participation will remain anonymous in perpetuity.

Your participation in the study is voluntary. Should you choose to withdraw from participation at any time you may do so without demur. If you choose to withdraw, your completed answers will be permanently deleted. The results of the study will be published as a dissertation, but your name will not be associated with any results.

This research poses no foreseeable risk to any of the participants in the study. All collected data will be kept confidential but, as with any research study, there always exists the very small risk that confidentiality will be compromised. Safeguards are in place to greatly minimize this risk. Although there may be no direct benefit to you to participate, the possible benefit of your participation may help by providing leaders in education statewide with information about how policy translates into practice with specific focus on how teacher evaluation has been implemented.

If you have any questions concerning the research study, you can contact me at (216) 375-8132.

Sincerely,

Caroline Johns

By signing this form, I acknowledge that I understand the nature of the study, the potential risks to me as a participant, and the means by which my identity will be kept confidential. My signature on this form also indicates that I am 18 years old or older, and that I give my permission to voluntarily serve as a participant in the study described.

Signature of participant \_\_\_\_\_ Date \_\_\_\_\_

Signature of researcher \_\_\_\_\_ Date \_\_\_\_\_



## **APPENDIX C**

### **ELECTRONIC SURVEY**

1. In your current position as principal, what grade levels does your building include? Please select the response that BEST describes your building.

- a.) Elementary (K-6)
- b.) Middle School (5-9)
- c.) High School (8-12)
- d.) Middle and High School (5-12)

2. How many years, including the current school year, have you been in your present position?

- a.) 1 year
- b.) 2 years
- c.) 3 or more years

(\*Respondents answering “A” were exited from the survey)

3. How many years have you been observing teachers using the Pennsylvania Department of Education Educator Effectiveness model (include the current school year)?

- a.) 2 or more years
- b.) This is the first year of implementation
- c.) Not currently using due to an existing Collective Bargaining Agreement

(\*Respondents answering “B” or “C” will be discontinued)

4. Which county is your school district located in?

**Participant Experience:**

5. Have you ever participated in training on Educator Effectiveness?

a.) Yes

b.) No

6. Please indicate the type of training you had on Educator Effectiveness and approximate length of time. If you did not participate in training, please mark “not applicable.”

	Not Applicable	1-2 Hours	Half Day	1 Day	2-3 Days	Training is on-going
Training provided by the Intermediate Unit						
Training provided by the school district						
Participation in a Professional Learning Community in my district or building						
Wed-based training (e.g. Teachscape)						

I have engaged in my own independent learning						
Other (Please explain)						

7. The four domains of Educator Effectiveness are based on the Charlotte Danielson model.

Were you familiar with the Danielson model prior to implementation of Educator Effectiveness?

a.) Yes

b.) No

8. When you were a teacher, were you observed using an evaluation tool based on the Danielson model?

a.) Yes

b.) No

9. Prior to implementation of Educator Effectiveness, did you ever observe teachers using an evaluation tool based on Danielson's four domains?

a.) Yes

b.) No

10. Please choose the statement that best describes how you would rate your level of preparedness with evaluating teachers using the Educator Effectiveness model.

a.) Very prepared

b.) Prepared

c.) Somewhat prepared

d.) Not prepared

11. Please indicate how effective each of these items were in helping prepare you to evaluate teachers using the Educator Effectiveness model.

	Extremely Effective	Very Effective	Moderately Effective	Slightly Effective	Not Effective At All	Not Applicable
Trainings provided by the Intermediate Unit						
Training provided by the school district						
Web-based training (e.g. Teachscape)						
Participation in a Professional Learning Community in your district						
Independent Learning Activities						
Prior experience evaluating teachers using a tool based on Danielson's four domains.						
Other (please explain)						

12. Based on your experience in your current building, indicate your level of agreement with the following statement, “The Educator Effectiveness evaluation model has helped ensure there is an effective teacher in every classroom.”

- a.) Strongly agree
- b.) Agree
- c.) Somewhat agree
- d.) Neither agree or nor disagree
- e.) Somewhat disagree
- f.) Disagree
- e. Strongly Disagree

13. When planning for professional development, does your district take into consideration feedback from teacher evaluations?

- a.) Yes
- b.) No
- c.) Not sure

14. Overall, how well do you think the teachers in your building understand the four domains that comprise the “observation/practice” portion of the rubric?

- a.) Extremely well
- b.) Very well
- c.) Moderately well
- d.) Slightly well
- e.) Not well at all

15. How many administrators typically observe each teacher in a school year?

a.) 1

b.) More than 1

c.) More than 1 but only if they are a struggling teacher

16. What do you think policymakers were hoping to accomplish by creating the teacher evaluation system that is now in place?

17. Is there anything else you would like to share about your experience with the Educator Effectiveness evaluation model?

18. Would you be willing to participate in a follow-up discussion with the researcher? If yes, please provide your contact information.

## APPENDIX D

### INTERVIEW PROTOCOL

Part I: Notes for the Interviewer
-----------------------------------

#### Overview

1. Tape-record the interviews if permission is granted
2. Interview in a neutral setting.
3. Each interview will be approximately 45-60 minutes.

#### Interview Methodology:

Interviews will be implemented with a customized approach allowing for an in-depth investigation. Follow-up questions (i.e. prompts) will be used to stimulate interviewee memory.

The interviewer will use a semi-structured question design (Part III). Interview contains:

1. A predetermined set of 17 questions
2. All predetermined questions were the same for respondents

Designation of Interviewee: \_\_\_\_\_

Location of Interview: \_\_\_\_\_

Date: \_\_\_\_\_

Start Time: \_\_\_\_\_

Finish Time: \_\_\_\_\_

## Part II: Components of the Interview

### **1. Components of the Interview**

- a. Introduction (5-10 minutes)
- b. Review confidentiality and consent form.
- c. Create a relaxed environment
- d. Dialogue

*Question:* Have you received my introductory correspondence explaining my research and the format that will be used?

*Question:* Are there any questions?

### **2. Explain the purpose of the interview**

Thank you for agreeing to talk with me. As I stated previously, I am completing a policy study of Act 82, specifically focusing on Educator Effectiveness and the resulting teacher evaluation tool that you are currently using. During the time we have together, I would like to get an understanding of your experiences and observations as they relate to this topic.

### **3. Ask permission to record interview**

With your authorization, I would like to tape-record our discussion to get an inclusive record of what is said, since the notes I take will not be as comprehensive as I will require. No one other than I will listen to anything you say to me. Only I will have access to the records. The research results will describe what you and others have said predominantly in summation. No responses will be ascribed to you by name.

The open-ended questions are intended to obtain your personal experience and perceptions. I anticipate that the interview time will take approximate 30 minutes. If you agree to volunteer and



participate in the research process, please sign the informed consent page and confidentially agreement.

Would you give me permission to tape the interview?

Do you have any questions before we begin?

Part III: Interview Questions
-------------------------------

Research Questions:

RQ2: What are school principals' understandings of the intended purposes of Act 82?

RQ3: How have principals been prepared for the implementation of the Act 82 Teacher Effectiveness policy?

RQ4: How have principals implemented the Act 82 Teacher Effectiveness policy in their buildings?

**Participant's Background Experience:**

**Interview Question 1:** Please tell me about your professional experience in education.

**Interview Question 2:** Based on what you remember, what was your experience like with teacher evaluation when you were a teacher? (RQ4)

*Prompts:* Would you say it was more similar or different than the current evaluation system? How so?

How would you say that the evaluation system you were observed with as a teacher compares to the one you are currently using?

**Interview Question 3:** During your time as an administrator, did you have an opportunity to evaluate teachers using a method other than what you are using now? If yes, can you please describe it and talk about how it compared to your current method of evaluation. (RQ 3&4)

**Interview Question 4:** How were you prepared or what kind of training did you receive in order to successfully implement the new evaluation system? (RQ3)

*Prompts:* How did the training help you become better prepared?

*Prompts:* What kind of training do you think could have helped you become better prepared?

*Prompts:* Has there been anything you've done on your own to help yourself with using this evaluation tool? This could include looking at resources such as the SAS website, reading Danielson materials, talking with colleagues who use this tool, etc.

**Interview Question 5:** Did you feel you were adequately prepared? (RQ3 & 4)

*Prompt:*

(If not adequately prepared...)

What do you think could have been done differently to better prepare you for implementation?

Do you feel comfortable using the tool now that you have more experience with it?

(If adequately prepared...)

Please tell me more about the training and experiences you've had that helped you feel prepared.

### **Participant's Present Experience:**

**Interview Question 6:** Please talk to me about your experiences with using the current evaluation system to actually observe and evaluate teachers. (RQ4)

**Interview Question 7:** Do you feel the new evaluation method has had an impact on the quality of instruction? (RQ4)

*Prompts:*

(If yes...) Can you tell me more about how you have seen instruction change, and why you think that change has occurred.

(if no...) One of the purposes of this current evaluation tool was to ensure high quality instruction in each classroom. Why do you think it's not had the intended impact?

**Interview Question 8:** In general, what's been the reaction of your teachers to having student data be a part of their evaluation? (RQ4)

*Prompt:*

How do you talk to them about data? Is that done during an evaluation meeting or at other times?

What types of student data do you and the teachers discuss in relation to a teacher's performance?

**Interview Question 9:** Are you the only one who observes teachers in your building or are they observed by other administrators? (RQ4)

*Prompt:*

(If multiple administrators in building) How do you decide who observes the teachers?

**Interview Question 10:** Can you share with me how your district goes about the process of planning for professional development? (RQ4)

*Prompts:* Who decides topics to cover?

How do they make those decisions?

**Interview Question 11:** Since the implementation of Act 82 in your district, have you had any additional professional development or support provided? Supports could include time to reflect with your colleagues, support provided by central administration, training at the intermediate unit, online training, etc. (RQ3)

**Interview Question 12:** Overall, what has the teacher response been to this evaluation tool? (RQ4)

*Prompts:*

Have teachers who've expressed an opinion been generally positive or negative about it?

Can you give me examples of some of the feedback they have provided?

**Reflection on Experience:**

**Interview Question 13:** What do you think policymakers were hoping to accomplish by creating the teacher evaluation system that is now in place? (RQ2)

*Prompts:*

Why do you think legislators have become involved with how teachers are evaluated?

We know that accountability in education has risen dramatically. How do you think teacher evaluation plays into accountability?

**Interview Question 14:** Was there anything in your past experience, training or evaluation tools you worked with, that helped prepare you for implementing this evaluation system? (RQ3)

**Interview Question 15:** Has the new evaluation system impacted your ability to remove ineffective teachers? (RQ 2&4)

*Prompts:*

What challenges, if any, do you think you would face now trying to remove an ineffective teacher?

Is this different than what you would have experienced prior to the passage of Act 82?

**Interview Question 16:** Have there been any unintended consequences, positive or negative, that have occurred as a result of the current evaluation method? (RQ4)

**Interview Question 17:** Is there anything else that you would like to share with me about your experience with teacher evaluation?

## **APPENDIX E**

### **FOLLOW-UP SCRIPT FOR A LOW RESPONSE RATE**

Hello, my name is Caroline Johns and I am a currently a student at the University of Pittsburgh completing a doctorate in Administrative and Policy Studies. I am conducting a research study entitled: *A Mixed Method Exploration Of The Efficacy Of The Teacher Evaluation System in Western Pennsylvania*. This is a policy study looking at how the intended purposes of Educator Effectiveness (Act 82) have been translated into practice. This study has been approved by the University of Pittsburgh Human Research Protection Office.

I am calling/speaking/emailing to you today to request your participation in this study by completing a short survey. I found your contact information on your district's website. Your participation in this research project is completely voluntary, and there are no known risks beyond those encountered in everyday life. In the survey, you will be asked about your experience with implementing Act 82 Educator Effectiveness as well as your thoughts about its effectiveness as a teacher evaluation tool. All information is kept confidential, and you may choose to exit the survey at any time. The survey takes approximately 10 minutes to complete.

The last question in the survey asks if you would be willing to participate in a follow up interview. This interview can occur in person or over the phone and will be scheduled at your convenience. The interview will take approximately thirty minutes. A short follow up interview

or phone call may be necessary for clarification of your comments if needed. Your participation will be kept confidential, your name will not appear on any document related to the study, and your participation will remain anonymous in perpetuity.

## BIBLIOGRAPHY

- America's Perfect Storm, (n.d.). In Educational testing service. Retrieved from [http://www.ets.org/perfect\\_storm/](http://www.ets.org/perfect_storm/).
- Anderson, J. E. (2011). *Public policymaking*. Boston, MA: Wadsworth Cengage Learning.
- Barbach, E. (1977). *The implementation game: What happens after a bill becomes law*. Cambridge, MA: MIT Press.
- Baumer, D. C., & Van Horn, C. E. (2014). *Politics and public policy: Strategic actors and policy domains*. Washington, DC: CQ Press.
- Baumgartner, F. R., & Jones, B. D. (1993). *Agendas and instability in American politics*. Chicago, IL: The University of Chicago Press.
- Bernard, S. C. & Mondale, S. (2001). *School. The story of American public education*. Boston, MA: Beacon Press.
- Bill & Melinda Gates Foundation. (2012). Measures of Effective Teaching (MET) Project. Seattle, WA: Retrieved from [www.metproject.org](http://www.metproject.org).
- Birkland, T. A. (2010). *An introduction to the policy process: Theories, concepts, and models of public policymaking*. New York, NY: M.E. Sharpe.
- Braun, H., Kirsch, I., Sum, A., & Yamamoto, K. (2007). *America's perfect storm: Three forces changing our nation's future*. Princeton, N.J., Educational Testing Service. Retrieved from [http://www.ets.org/Media/Education\\_Topics/pdf/AmericasPerfectStorm.pdf](http://www.ets.org/Media/Education_Topics/pdf/AmericasPerfectStorm.pdf)
- Brinkman, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing*. Thousand Oaks, CA: Sage Publications, Ltd.
- Cahn, M. A. (1995). The players: Institutional and non-institutional actors in the policy process. In Theodoulou, S.Z. & Cahn, M. (Ed.), *Public policy: The essential readings* (pp. 201-211). Upper Saddle River, NJ: Prentice Hall, Inc.
- Charmaz, K. (2014). *Construction grounded theory*. Thousand Oaks, CA: Sage Publications, Ltd.

- Cibulka, J. G. (1995). Policy analysis and the study of the politics of education. In J. D. Scribner & D. Layton (Eds.), *The study of educational politics* (pp. 105-125). London, UK: Falmer Press.
- Clemons, R. S., & McBeth, M. K. (2008). *Public policy praxis*. New York, NY: Pearson Longman.
- Cobb, R. W., & Elder, C. D. (1972). *Participation in American politics: The dynamics of agenda building*. Baltimore, ME: Johns Hopkins University Press.
- Cochran, C. E., Mayer, L. C., Carr, T. R., Cayer, N. J., McKenzie, M. J., & Peck, L. (2012). *American public policy: An introduction*. Boston, MA: Wadsworth Cengage Learning.
- Coffman, C., Gonzalez-Molina, G., & Clifton, J. K. (2002). *Following this path: How the world's greatest organizations drive growth by unleashing human potential*. New York: Warner Books.
- Cohen, D. K., Moffitt, S. L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education*, 113 (4), pp. 515-548.
- Corey, S.M. (1954). Action research in education. *The Journal of Educational Research*, 47(5), pp. 375-380.
- Creswell, J. W. (2012). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: ASCD.
- Danielson, C. (2016). *Talk About Teaching!: Leading Professional Conversations*. Thousand Oaks, CA: Sage Publications.
- Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York, NY: Teachers College Press.
- DeLeon, P. (1988). *Advice and consent: The development of the policy sciences*. New York, NY: Russell Sage Foundation.
- DeLeon, P. (1999). The stages approach to policy process: What has it done? Where is it going? In Sabatier, P.A. (Ed.), *Theories of the policy process* (pp. 19-32). Boulder, CO: Westview Press.
- Department of Education. (2011). Department of Education Awards \$200 Million to Seven States to Advance K-12 Reform. Retrieved from [www.ed.gov](http://www.ed.gov).



- Domhoff, G. W. (2013). *Who rules America? The triumph of the corporate rich*. Boston, MA: McGraw Hill.
- Dominick, J. R. (1984). Education as portrayed in network TV newscasts. *Educational Horizons*, 62(3), 86-89.
- DuFour, R., Eaker, R. & DuFour, R. (2005). *On common ground: The power of professional learning communities*. Bloomington, IN: Solution Tree.
- Edelman, M. (1977). *Political language: Words that succeed and policy that fail*. New York, NY: Academic Press.
- Editorial Projects in Education Research Center. (2004, August 3). Issues A-Z: Accountability. *Education Week*. Retrieved January 15, 2017 from <http://www.edweek.org/ew/issues/accountability/>
- Elder, C. D., & Cobb, R. W. (1983). *The political uses of symbols*. New York, NY: Longman, Inc.
- Fairbanks, C.M. & LaGrone, D. (2006). Learning together: constructing knowledge in a Teacher Research Group. *Teacher Education Quarterly*, 33 (3), pp. 7-25.
- Fesler, J. W. (1983). Politics, policy, and bureaucracy at the top. *Annals of the American Academy of Political and Social Science*, 466, pp. 23-41.
- Flemming, A. S. (1960). The philosophy and objectives of the National Defense Education Act. *Annals of the American Academy of Political and Social Science*, 327, pp. 132-138.
- Fowler, F. C. (2013). *Policy studies for educational leaders*. Upper Saddle River, NJ: Pearson Education, Inc.
- Fuhrman, S., Firestone, W., & Kirst, M. (1989). *Progress of reform: An appraisal of state education initiatives*. Washington, DC: Center for Policy Research in Education.
- Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F. & Kwang, S.Y. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), pp. 915-945.
- Glanz, J. (2005). Action research as instructional supervision: Suggestions for principals. *NASSP Bulletin*, 89 (17), pp. 17-27.
- Gordon, R., Kane, T .J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: The Brookings Institution.
- Graber, D. (1988). *Processing the news: How people tame the information tide*. New York, NY: Longman.

- Grossberg, M. (2011). From feeble-minded to mentally retarded: Child protection and the changing place of disabled children in the mid-twentieth century United States. *Pedagogical Historica*, 47(6), 729-747.
- Guest, G., Bunce, A. & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.
- Guthrie, J. W. (1982). The future of federal education policy. *Education and Urban Society*, 14(4), 511-530.
- Hagevik, R., Aydeniz, M. & Rowell, C.G. (2012). Using action research in middle level teacher education to evaluate and deepen reflective practice. *Teaching and Teacher Education*, 28, pp. 675-684.
- Hamilton, L .S., Stecher, B .M., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J. L., Naftel, S., & Barney, H. (2007). *Standards-Based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND Corporation.
- Hanushek, E. (1996). School resources and student performance. In Burtless, G. (Ed.), *Does money matter? The effect of school resources on student achievement and adult success*. Washington, D.C.: Brookings.
- Harris, A. & Jones, M. (2010). Professional learning communities and system improvement. *Improving Schools*, 13, pp. 172-181.
- Heck, R. H. (2004). *Studying educational and social policy: Theoretical concepts and research methods*. New York, NY: Routledge Taylor & Francis Group.
- Heibert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education*, 30(1), 3-19.
- Hess, F. M. (2003). Refining or retreating? High-stakes accountability in the states. In Peterson, P. E. & West, M. R. (Ed.), *No child left behind? The politics and practice of school accountability* (pp. 1-20). Washington, D.C.: Brookings Institution.
- Hochschild, J. L. (2009). *Conducting intensive interviews and elite interviews*. Workshop on interdisciplinary standards for systemic qualitative interviews.
- Honig, M. I. (2006). New directions in education policy implementation. Albany, NY: State University of New York Press.
- Huck, S. W. (2012). Reading statistics and research. Boston, MA: Pearson.
- Hulley, S. B. (2007). *Designing clinical research* (3<sup>rd</sup> ed.). Philadelphia, PA: Lippincott.

- Hunt, S. L., & Staton, A. Q. (1996). The communication of educational reform: A nation at risk. *Communication Education*, 45, 271-292.
- Iyengar, S., & Kinder, D. (1987). *News that matters: Television and American opinion*. Chicago, IL: University Press of Chicago.
- Kaestle, C.F. (2001). The common school. In Mondale, S. & Patton, S.B. (Ed.), *School: The story of American education* (pp. 11-17). Boston: Beacon Press.
- Kane, T. J., & Staiger, D. O. (2003). Unintended consequences of racial subgroup rules. In Kernell, S., Jacobson, G. & Kousser, T. (2011). *The logic of American politics*. Washington, DC: CQ Press.
- Kingdon, J. W. (2011). *Agendas, alternatives and public policies*. Glenview, IL: Pearson.
- Kirst, M. W., & Wirt, F. M. (2009). *The political dynamics of American education*. Richmond, CA: McCutchan Publishing Corporation.
- Kiser, L. L., & Ostrom, E. (1982). The three worlds of action: A metatheoretical synthesis of approaches. In Ostrom, E. (Ed.), *Strategies of political inquiry* (pp. 56-88). Beverly Hills, CA: Sage.
- Knowles, M.S., Hotlon, E.F., & Swanson, R.A. (2012). *The adult learner: The definitive classic in adult education and human resource development*. New York: Routledge.
- Kraft, M., & Furlong, S. (2010). *Public policy: Politics, analysis and alternatives*. Washington, DC: CQ Press.
- Kruschke, E. R., & Jackson, M. B. (1987). *The public policy dictionary*. Santa Barbara, CA: ABC-CLIO.
- Learning Forward (2013). Standards for professional learning. Retrieved from [www.learningforward.org](http://www.learningforward.org).
- Lee, J.F.K. (2008). A Hong Kong case of lesson study-Benefits and concerns. *Teaching and Teacher Education*, 24, pp. 1115-1124.
- Lewis, C. (2002). *Lesson Study: A Handbook for Teacher-Led Improvement Instruction*. Philadelphia, PA: Research for Better Schools.
- Linn, R. L., Baker, E. L. & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3-16.
- Lolly, J. L. (2009). The National Defense Education Act, current STEM initiative, and the gifted. *Gifted Child Today*, 32(2), p. 50.

- Majone, G. (1989). *Evidence, argument, and persuasion in the policy process*. New Haven, CT: Yale University Press.
- Mantzoukas, S. (2004). Pearls, pith, and provocation: Issues of representation within qualitative inquiry. *Qualitative Health Research*, 14(7), pp. 994-1004.
- Martin, E. W., Martin, R. & Terman, D. L. (1996). The legislative and litigation history of special education. *Special Education for Students With Disabilities*, 6(1). Retrieved from [www.futureofchildren.org](http://www.futureofchildren.org).
- Mead, J. F. (2009). The role of law in educational policy formation, implementation and research. In Sykes, G., Schneider, B. & Plank, N. (Ed.), *Handbook of education policy research* (pp. 286-295). New York, NY: American Educational Research Association.
- McBride, A. (2006, December). *Brown v. Board of Education*. Retrieved from [www.pbs.org](http://www.pbs.org).
- McDonnell, L M., & Elmore, R. F. (1987). Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis*, 9, pp. 133-152.
- McGuinn, P. J. (2006). *No child left behind and the transformation of federal education policy, 1965-2005*. Lawrence, KS: University Press of Kansas.
- McGuinn, P. J. (2014). Presidential policymaking: Race to the top, executive power, and the Obama education agenda. *The Forum*, 12(1), pp. 61-79.
- Mills, C. W., & Wolfe, A. (2000). *The power elite*. New York, NY: Oxford University Press, Inc.
- Mitchell, W. (1962). *The American polity*. New York, NY: Free Press.
- National Association of Elementary Principals (2008). *Leading learning communities: Standards for what principals should know and be able to do*. Alexandria, VA: NAESP.
- National Association of State Directors of Special Education. (1994). *Summary of goals 2000: Educate America act*. Alexandria, VA. Retrieved from [files.eric.ed.gov/fulltext/ED389130.pdf](http://files.eric.ed.gov/fulltext/ED389130.pdf).
- National Commission of Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Ann Arbor, MI: University of Michigan Library.
- National Commission on Teaching & America's Future. (1996). *What matter's most: Teaching for America's future*. Retrieved from <http://www.namodemello.com.br/pdf/tendencias/whatmattersmost.pdf>.
- Neustadt, R. E. (1990). *Presidential power and the modern presidents*. New York, NY: The Free Press.

- NVivo9. (2011). *Software for qualitative analysis*. Retrieved from [http://www.qsrinternational.com#tab\\_yo](http://www.qsrinternational.com#tab_yo)
- Odden, A. R. (1991). The evolution of education policy implementation. In Odden, A.R. (Ed.), *Education implementation policy* (pp. 1-12). Albany, NY: State University of New York Press.
- Ostrom, E., Cox, M., & Schlager, E. (2014). An assessment of the institutional analysis and development framework and introductions of the socio-ecological systems framework. In Sabatier, P. & Weible, C. M. (Ed.), *Theories of the policy process* (pp. 267-306). Boulder, CO: Westview Press.
- Pennsylvania Department of Education. (2014). *Educator effectiveness administrative manual*. Harrisburg, PA.
- Pennsylvania House Bill 1901, Act 82 (2012).
- Pennsylvania School Performance Profile. (2016). Retrieved from [www.paschoolperformanceprofile.org](http://www.paschoolperformanceprofile.org).
- Perry, R. R., & Lewis, C. C. (2009). What is successful adaptation of lesson study in the US? *Journal of Educational Change*, 10(4), pp. 365-391.
- Peterson, P. E., & West, M. R. (Ed.). *No child left behind? The politics and practice of school accountability*. Washington, D.C.: The Brookings Institute.
- Poliski, M. M., & Ostrom, E. (1999). *An Institutional framework for policy analysis and design*. (Working Paper No. W98-27). Retrieved from the George Mason University website: <http://mason.gmu.edu/~mpolski/publications.html>.
- Porter, A. C., Garet, M. S., Desimone, L., Yoon, K. S., & Birman, B. F. (2000). Does professional development change teaching practice? Results from a three-year study. Report prepared for U.S. Department of Education, Washington, D.C.
- Ravitch, D. (2001). As American as public school. In Mondale, S. & Patton, S.B. (Ed.), *School: The story of American education* (pp. 63-70). Boston: Beacon Press.
- Razfar, A. (2011). Action research in urban schools: empowerment, transformation, and challenges. *Teacher Education Quarterly*, 38 (4), pp. 25-44.
- Reeves, D. (2004). *Accountability for Learning: How Teachers and School Leaders Can Take Charge*. Alexandria, VA: ASCD.
- Rudalevige, A. (2003). No child left behind: Forging a congressional compromise. In Peterson, P. E. & West, M. R. (Ed.), *No child left behind? The politics and practice of school accountability* (pp. 1-20). Washington, DC: Brookings Institution.

- Sabatier, P. A. (2007). *Theories of the policy process* (2nd ed.). Cambridge, MA: Westview Press.
- Sabatier, P. A., & Weible, C. M. (2007). The advocacy coalition framework: Innovations and clarifications. In Sabatier, P. A. (Ed.), *The policy process* (pp. 189-220). Boulder, CO: Westview Press.
- Sarason, S. B., & Doris, J. (1982). Public policy and the handicapped: The case of mainstreaming. In Lieberman, A. & McLaughlin, M.W. (Ed.), *Policy making in education* (pp.23-55). Chicago, IL: The University of Chicago Press.
- Schattschneider, E. E. (1975). *The semi-sovereign people*. Boston, MA: Wadsworth, Cengage Learning.
- Schwartz, R. B., & Robinson, M. A. (2000). Goals 2000 and the standards movement. *Brookings Papers on Education Policy*, 3, 173-214.
- Seidman, I. E. (1991). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York, NY: Teachers College Press.
- Spillane, J. P. (2000). Cognition and policy implementation: District policy-makers and the reform of mathematics education. *Cognition and Instruction*, 18 (2), pp. 141-179.
- Spring, J. (2014). *Political agendas for education: From race to the top to saving the planet*. New York, NY: Routledge.
- Sroufe, G. E. (1995). Politics of education at the federal level. In Scribner, J. D. & Layton, D. H (Ed.), *The study of educational politics* (pp. 75-88). New York. NY: The Falmer Press.
- Stedman, J. B. (1994). Goals 2000: *Overview and analysis*. Congressional Research Service, 144.
- Steigler, J. W. & Hiebert, J. (2000). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: The Free Press.
- Stepanek, J., Appel, G., Leong, M., Mangan, M. T., & Mitchell, M. (2007). *Leading Lesson Study: A Practical Guide for Teachers and Facilitators*. Thousand Oaks, CA: Corwin Press.
- Stout, R. T., Tallerico, M., & Scribner, K. P. (1995). Values: the 'What?' of the politics of education. In Scribner, J.D. & Layton, D.H. (Ed.), *The study of educational politics*. New York, NY: The Falmer Press.
- Stromberg, D., & Snyder, Jr., J. M. (2008). The media's influence on public policy decisions. In Roumeen, I. (Ed.), *Information and public choice: From media markets to policymaking*. Herndon, VA: The World Bank.

- Sunstein, B. S., & Chiseri-Strater, E. (2007). *Fieldworking: Reading and writing research 3rd Edition*. Boston, MA: Bedford/St. Martin's Press.
- Task Force on Teaching as a Profession. (1986). *A nation prepared: Teachers for the 21<sup>st</sup> Century*. New York, NY: Carnegie Forum on Education and the Economy.
- Theodoulou, S. Z. (1995). How public policy is made. In Theodoulou, S. Z. & Cahn, M. (Ed.), *Public policy: The essential readings* (pp. 86-95). Upper Saddle River, NJ: Prentice Hall, Inc.
- Theodoulou, S. Z., & Cahn, M. (1995). *Public policy: The essential readings*. Upper Saddle River, NJ: Prentice Hall, Inc.
- True, J. L., Jones, B. D., & Baumgartner, F. R. (2007). Punctuated equilibrium theory: Explaining stability and change in public policymaking. In Sabatier, P. A. (Ed.), *Theories of the policy process* (pp. 155-187). Boulder, CO: Westview Press.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- United States Department of Education. (2011). Department of Education Awards \$200 Million to Seven States to Advance K-12 Reform. Retrieved from [www.ed.gov](http://www.ed.gov).
- United States Department of Education. (2009). Race to the Top Program Executive Summary. Retrieved from [www2.ed.gov](http://www2.ed.gov).
- United States Department of Agriculture. (2016). Retrieved from <http://www.fns.usda.gov/school-meals/income-eligibility-guidelines>.
- Walker, J. L. (1991). *Mobilizing interest groups in America: Patrons, professions, and social movements*. Ann Arbor, MI; University of Michigan Press.
- West, M. R., & Peterson, P. E. (2003). The politics and practice of accountability. In Peterson, P. E. & West, M. R. (Ed.), *No child left behind? The politics and practice of school accountability* (pp. 1-20). Washington, D.C: Brookings Institution.
- Yarbrough, D. B., Shulha, L. M., Hopson, R .K. & Caruthers, F .A. (2011). *The program evaluation standards*. Thousand Oaks, CA: Sage.
- Yin, R. (2011). *Case study research design and methods* (3rd. ed.). Thousand Oaks, CA: Sage Publications.
- Zepeda, S. J. (2012). *Professional development: What works*. Larchmont, NY: Eye on Education.