

Spring 2008

# Teachers' Attitudes Toward Teacher Evaluation In High and Low Achieving Middle Schools As Measured By The Pennsylvania State System of Assessment

Janell Logue-Belden

Follow this and additional works at: <https://dsc.duq.edu/etd>

---

## Recommended Citation

Logue-Belden, J. (2008). Teachers' Attitudes Toward Teacher Evaluation In High and Low Achieving Middle Schools As Measured By The Pennsylvania State System of Assessment (Doctoral dissertation, Duquesne University). Retrieved from <https://dsc.duq.edu/etd/833>

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact [phillips@duq.edu](mailto:phillips@duq.edu).

TEACHERS' ATTITUDES TOWARD TEACHER EVALUATION IN HIGH-  
AND LOW-ACHIEVING MIDDLE SCHOOLS AS MEASURED BY THE  
PENNSYLVANIA STATE SYSTEM OF ASSESSMENT

A Dissertation

Submitted to the Interdisciplinary Doctoral Program  
for Educational Leaders (IDPEL)

Duquesne University

Submitted in partial fulfillment of the requirements for  
the degree of Doctor of Educational Leadership

By

Janell Logue-Belden

May 2008

Copyright by  
Janell Logue-Belden

2008

**DUQUESNE UNIVERSITY**  
**SCHOOL OF EDUCATION**  
**INTERDISCIPLINARY DOCTORAL PROGRAM FOR**  
**EDUCATIONAL LEADERS**

***Dissertation***

Submitted in Partial Fulfillment of the Requirements  
For the Degree of Doctor of Education (Ed.D.)

***Presented by:***

**Janell Logue-Belden**  
B.A., French, Allegheny College, 1990  
M.A., Education, Allegheny College, 1991  
M.S., School Administration, Duquesne University, 1998

***November 27, 2007***

**TEACHER'S ATTITUDES TOWARD TEACHER EVALUATION IN HIGH- AND  
LOW-ACHIEVING MIDDLE SCHOOLS AS MEASURED BY THE  
PENNSYLVANIA STATE SYSTEM OF ASSESSMENT**

***Approved by:***

\_\_\_\_\_, Chair  
**Wilton L. Barber, Ph.D.**  
Retired Professor, Department of Educational Foundations & Leadership  
Duquesne University

\_\_\_\_\_, Member  
**Gibbs Y. Kanyongo, Ph.D.**  
Assistant Professor, Department of Educational Foundations & Leadership  
Duquesne University

\_\_\_\_\_, Member  
**Maureen Ungarean, Ed.D.**  
Principal, Sto-Rox Elementary School

## ABSTRACT

### TEACHERS' ATTITUDES TOWARD TEACHER EVALUATION IN HIGH- AND LOW-ACHIEVING MIDDLE SCHOOLS AS MEASURED BY THE PENNSYLVANIA STATE SYSTEM OF ASSESSMENT

By

Janell Logue-Belden

May 2008

Dissertation Supervised by Dr. Wil Barber

With the emergence of the Pennsylvania Academic Standards in 1998, and the No Child Left Behind Act of 2001, greater emphasis has been placed on teacher accountability. As part of this reform, teacher evaluation has been an item of particular importance. Research demonstrates that teacher evaluation has been prevalent since the colonial times; however, evaluation systems differ and lack consistency.

The purpose of this study was to understand the relationship between Teachers' Attitudes Toward Teacher Evaluation In High and Low Achieving Middle Schools Based on The Pennsylvania State System of Assessment. Several Educational Researchers (Darling-Hammond, 1986; Weber, 1987; Frels, Cooper & Reagan, 1996) have shown evidence that the effectiveness of teacher evaluation can be significantly enhanced when teachers and principals view similarly the aspects of evaluation.

This study attempted to determine if a link existed between higher PSSA scores and teacher attitudes toward their teacher evaluations.

Overall findings of this study offer implications for professional educational practice.

This study's data supported the conclusions of the bulk of scholarly literature, which demonstrated that most teacher evaluation systems did not assist teachers to grow professionally (McGreal, 1983; Sergiovanni & Starrat, 1988; Shinkfield & Stuffelbeam, 1995).

## TABLE OF CONTENTS

	Page
CHAPTER ONE: INTRODUCTION.....	1
Statement of the Problem.....	1
<i>School Reform</i> .....	1
<i>Teacher Evaluation and Student Achievement</i> .....	6
<i>Process of Evaluation</i> .....	9
<i>Problems of Evaluation</i> .....	11
Summary.....	18
Definition of Terms .....	20
CHAPTER TWO: REVIEW OF THE LITERATURE .....	24
History of Evaluation.....	24
Purpose of Evaluation.....	30
Problems of Evaluation.....	33
Common Methods of Teacher Evaluation .....	38
Summary.....	41
CHAPTER THREE: METHODOLOGY .....	43
Purpose of the study .....	43
Research Questions .....	44
<i>Research Hypotheses</i> .....	45
<i>Study Variables</i> .....	46

<i>Method of Data Collection</i> .....	47
<i>Instrument Development</i> .....	47
<i>Data Analysis</i> .....	49
CHAPTER FOUR: RESULTS .....	51
<i>Research Question One</i> .....	52
<i>Research Question Two</i> .....	55
<i>Research Question Three</i> .....	57
<i>Research Question Four</i> .....	60
<i>Research Question Five</i> .....	61
<i>Research Question Six</i> .....	62
Summary .....	64
CHAPTER FIVE: DISCUSSION.....	65
Summary of Findings .....	65
Conclusions .....	68
Recommendations for Further Research .....	71
Summary .....	72
References.....	74
Appendices.....	86
<i>Appendix A</i> .....	87
<i>Appendix B</i> .....	88
<i>Appendix C</i> .....	93
<i>Appendix D</i> .....	94

# CHAPTER ONE: INTRODUCTION

## Statement of the Problem

### *School Reform*

With the emergence of the Pennsylvania Academic Standards in 1998, and the No Child Left Behind (NCLB) Act of 2001, greater emphasis has been placed on teacher accountability. As part of this reform, teacher evaluation has been an item of particular importance. Key provisions of the No Child Left Behind Act establish that the quality of the school staff has a direct correlation to student academic performance, and that teacher quality does have a tremendous influence on student achievement, a primary factor in determining the success of schools (U.S. Department of Education, 2004). The supervision and evaluation of teachers are key tasks proving increasingly important under the mandates of NCLB. NCLB also requires that states become more involved in the evaluation process. The Pennsylvania Department of Education (PDE), for instance, has mandated several evaluation forms (PDE 426, 427, and 428) and procedures for teacher performance assessment, requiring evaluation in the areas of planning and preparation, classroom environment, instructional delivery, and professionalism. The forms used by the Commonwealth of Pennsylvania were based on the Danielson (1996) model, a standards-based evaluation model that identifies set standards that can be used to assess teaching performance. The standards focus on key “frameworks” in the professional practice of teaching.

The Framework for Teaching evaluation model identifies twenty-two components that define the role of a teacher. Multiple elements associated with each component define specific teaching knowledge and skills. These components and their elements are clustered into four main domains of teaching responsibility: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities (Danielson, 1996). Specific rubrics define a level of performance for each individual element within the twenty-two components. The compilation of these ratings present a portrait of a teacher's effectiveness.

For several decades, public education has come under severe scrutiny beginning with *A Nation at Risk: The Imperative for Educational Reform* in 1983 by the National Commission on Excellence in Education. This document asserted that schools failed to teach students the basic skills of reading, and that teachers needed to have more input on textbook decisions, instructional strategies, and curriculum design. The commission also recommended that:

Persons preparing to teach should be required to meet high educational standards to demonstrate an aptitude for teaching and to demonstrate competence in an academic discipline. Salaries for the teaching profession should be increased and should be professionally competitive, market sensitive, and performance based. Salary, promotion, tenure, and retention decisions should be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average teachers encouraged, and poor ones either improved or terminated (p. 30).

Following the release of *A Nation at Risk*, President Ronald Reagan began to emphasize merit pay increases for teachers. This, in turn, highlighted the commission's recommendations and the need for a careful examination of teacher evaluation practices. A report published in June 1983 labeled *The Task Force on Education for Economic Growth*, written by the Education Commission of the United States, mirrored many of the recommendations of *A Nation at Risk*:

We recommend that boards of education and higher education in each state, in cooperation with teachers and school administrators, put in place, as soon as possible, systems for fairly and objectively measuring the effectiveness of teachers and rewarding outstanding performances. We strongly recommend that states examine and tighten their procedures for selecting not only those who come into teaching, but also those who ultimately stay. Ineffective teachers, those who fall short repeatedly in fair and objective evaluations, should, in due course and with due process, be dismissed (p. 39).

President George H. W. Bush proposed America 2000 in response to the public's negative reaction about the condition of the education system reported by the previously noted commissions. Public opinion revealed a general belief that improving student achievement depends upon the efforts of parents, teachers, schools, and society (McLaughlin & Shepard, 1995). In March 1994, the Educate America Act, also referred to as Goals 2000, was signed into law. Goals 2000 awarded grants to participating states and districts to support communities in the development and implementation of their own standards-based education reforms. President William Jefferson Clinton signed Goals 2000 in reaction to the public's strong cries for school reform. Clinton labeled this

initiative the “most grassroots oriented program ever promoted for improving the quality of education for our children” (Missing Citation). The endorsement of Goals 2000 was based on several principles that were thought to effect school change such as:

1. All students can learn;
2. Lasting improvements depend on school-based leadership;
3. Simultaneous top-down and bottom-up reform is necessary;
4. Strategies must be locally developed, comprehensive, and coordinated;
5. The whole community must be involved in developing strategies for system-wide improvement (U.S. Congress, 1993).

NCLB has received considerable media exposure since its inception. Teacher quality and efficacy has come under greater scrutiny as politicians and the public have a catchphrase on which to focus. Both presidential candidates in the 2004 election campaigned for educational reform, citing NCLB and serving to further impact the public with concerns for America’s schools. Local and national papers publish student scores and teacher salaries while the media continues to ask the question “What is wrong with our schools?” With many schools placed on the state school-improvement list, districts are challenged to address the public outcry concerning teacher accountability.

Administrators and state officials have recently put more focus on teacher evaluation and methods used to assess it. This focus has resulted in more inherent problems such as time, organization, and attitudes toward evaluation.

Overburdened principals who visit classrooms only a few times per year typically carry out teacher evaluations. Typically, teachers are asked to sign off on the forms and are not provided adequate time to consult with their supervisors to discuss pedagogy,

professional development, or new trends in educational outcomes. This lack of consistent mentoring stems not through a fault in either party, but from time constraints, creating frustration for all those involved. Teachers tend to resent a process that they perceive to be punitive in nature or a waste of their time (McGreal, 1983; McGreal, 1988; Darling-Hammond, 1986). Principals try to balance organizational requirements and mandates of NCLB with maintaining school discipline and positive staff morale (McGreal, 1988; Duke, 1990). Some teachers are more knowledgeable in content area than their supervisors. This fact can undermine the whole evaluation process and make the teachers feel little value. Teachers do not learn from the process, and little incentive exists in the process to encourage their learning. Most teachers perform a canned lesson for the observation, one that meets the required checklist (Danielson & McGreal, 2000).

Not until recently has there been a move to change the way the evaluation process occurs in school districts. It would appear that for teacher evaluation to be effective and meet the demands of the twenty first century, it must consist of a process with shared goals and shared development by all participants. Additional consideration must be given to the obstacles to effective evaluation: attitudes of teachers and principals about each other, consistency in evaluation, evaluative feedback, and multiple evaluative purposes (Danielson, 2000).

Research has shown that what teachers know and do is one of the most important factors in determining student achievement (Darling-Hammond, 1997). In *What Matters Most: Teaching for America's Future* (1996), the National Commission on Teaching and America's Future asserted,

What teachers know and can do is the most important influence on what students learn; recruiting, preparing, and retaining good teachers is the central strategy for improving schools; and school reform cannot succeed unless it focuses on creating the conditions in which teachers can teach, and teach well (p. 10).

A clear link exists between teacher effectiveness and student achievement.

Glasser (1985) illustrates this link in conclusions where he explains the focus of classroom learning is much more the responsibility of the teacher than that of the student. Sergiovanni (1991) supported this claim in his assertion that the teacher is the one factor within the classroom setting that makes the critical difference. Teacher evaluations are important because they have consistently been linked to higher student achievement.

#### *Teacher Evaluation and Student Achievement*

Teachers have substantial impact on student learning; in fact, teacher classroom practices are likely to be an important pathway for these effects. According to Odden and Kelly (2002) in *Paying Teachers for What They Know and Do*, scores from well designed, performance based teacher evaluation systems may provide a measure of important teacher behaviors that can be used in a comprehensive model of teacher classroom and school effects on student achievement.

Teacher evaluation scores prove useful as representations of teaching practices affecting student learning positively. In a study on the relationship between teacher performance evaluation scores and student achievement, Milanowski, Kimball, and White (2004) examined the relationship between teacher evaluation scores in a Cincinnati school district with a rigorous standards based teacher evaluation system and a value-added measure of student achievement. Findings revealed that teacher evaluation scores

had a moderate degree of criterion-related validity. The teacher assessment system was able to identify which teachers had students with higher than expected levels of achievement as measured by test scores, to a degree greater than chance. The empirical results showed that evaluations produced by a relatively rigorous standards based system are related to an accepted measure of student learning. Given favorable evidence on the construct validity of teacher evaluation scores as representations of teaching behavior, it was concluded that these scores may be a useful part of the framework for assessing.

Clear differences exist in the strength of the relationship between teacher evaluation scores and classroom achievement in reading compared with math or language arts. Empirical research completed by Gallagher (2002) measured the relationship between measures of teacher quality and student achievement in an elementary school in California. The researcher estimated classroom effects and analyzed their relationship to teacher evaluation scores. The implications of the study showed that it is possible to improve teacher evaluation systems by specifying a clear definition of high quality teaching in specific subject areas and evaluating teachers based on that definition.

Similarly, Spiggle (2003) investigated the relationship between the teaching performance of 5<sup>th</sup> grade teachers and student achievement growth in 5<sup>th</sup> grade reading and mathematics in North Carolina Public Schools. The scores of the students were from the end of year tests, and the teacher performance was appraised with the North Carolina Performance Instrument that included five functions of teaching characteristics. Spiggle's research showed that the best overall predictor of students' mathematics improvement related to the results of the 3<sup>rd</sup> domain of the teacher characteristic—

instructional presentation. This model accounts for 23% of the variance in students' improved class average achievement in math and reading.

States that implement teacher evaluation scores achieve better student results on standardized tests and may be better able to reduce the achievement gaps between students from different social backgrounds. According to Danielson and McGreal (2000), four hundred teachers were rated on four composite scores in the areas of the Danielson Framework for Teaching while seven thousand students from grades 4-6 were reviewed using their Terra Nova scores. Using the standards-based teacher evaluation score as a measure of teacher quality, researchers found that teacher quality does affect student achievement and that teacher quality is not distributed equitably among classrooms with varying baseline achievement, poverty, and minority concentrations. Students from poor, minority, and low-achieving backgrounds have access to teachers of lower quality as reflected in the teacher evaluation scores.

Studies from Texas also give credence to the theory that teachers do have an impact on student achievement. Hutto (2001) examined the relationship between scores received by teachers on a new Texas Teacher Evaluation System and student achievement. This study identified a specific domain of the evaluation tool measuring improvement of academic performance for all students on campus. A comparison was completed using the student achievement scores as measured by the Texas Assessment of Academic Skills. The criteria for the teacher evaluation domain require that teachers become familiar with every student's prior performance. The assumption is that teachers who have personal knowledge of each student are more attentive to the differences; therefore, customize the lesson to ensure that all students succeed. Findings of the

research revealed a significant increase in student math and reading scores after implementation of the new teacher evaluation system. Grade level had little influence on achievement scores; in fact, findings showed that the greatest influence on student achievement included alignment of instruction, appropriate sequence of instruction, appropriate materials, monitoring of student performance and attendance, interaction with students in high-risk situations, implementation of an intervention plan, and provision of a campus-wide action program.

### *Process of Evaluation*

Virtually every public school district by order of state law or regulation has a formal procedure in place for the evaluation of teachers (Danielson & McGreal, 2000). Two models widely used are formative and summative evaluations. Summative evaluation focuses on accountability while formative evaluation focuses on professional growth. Summative evaluation practices are a fundamental part of every teacher evaluation system (Duke & Stiggins, 1986). The summative evaluation process collects data used for making decisions about a person's standing in an organization. Summative evaluations provide feedback and inform retention decisions including tenure and, in some districts, merit pay (Catano & Stronge, 2006). Direct observation of teachers is the center of this teacher evaluation process. Observations are typically conducted twice per year, with more frequent observations of non-tenured teachers or teachers found to be performing below expectations. The observation usually encompasses one class period, followed by the supervisor's formal documentation of the lesson. This documentation takes the form of a checklist or a narrative based on the district's policy.

Formative evaluation is designed to help teachers improve their instructional performance. Formative models concentrate on the professional development of teachers, promoting professional growth. The principle of formative evaluation is that teachers possess the ability to improve (Stiggins & Bridgeford, 1984). In essence, formative evaluation involves three features: it is ongoing, involves formative assessment, and seeks in-depth information about a teacher's practices (Beyer, 1995). The keystone of the formative evaluation process is the teacher who drives the improvement-oriented, goal-setting process involving collaboration between teachers and administrators (Darling-Hammond, Wise & Pease, 1983; Duke, 1990; Haefele, 1992).

The teacher evaluation process in numerous school districts has become a routine requirement, which does not increase the teachers' instructional strategies to enhance student learning (Danielson & McGreal, 2000). The process of teacher evaluation differs by varying degrees from district to district, and building to building. In some districts, teacher evaluations are just a matter of spending a few minutes in a classroom to observe if the teacher can manage a classroom and simply keep students in their seats. In other districts, principals may never visit a classroom, but base a rating on how well teachers submit their paperwork or how many phone calls are received from disgruntled parents. In still other districts, principals follow strict procedures and perform several observations (Catano & Stronge, 2006).

The teacher evaluation process has been a controversial issue over the years due to the fact that the standards used to evaluate teachers have continued to change throughout the history of teacher evaluation. Historically, the methods used for teacher evaluation have not been reflective of recent research, which empowers teachers to

become leaders in their classrooms by encouraging them to design goals and reflect on their teaching techniques (Danielson & McGreal, 2000).

Based on research, it is evident that teacher effectiveness is an essential factor that has a profound impact on student learning. Teacher quality is the single most important determinant of student achievement (Goldhaber, Brewer, & Anderson, 1999). It is also noted that teacher knowledge of subject matter and teacher qualifications, including education and experience, account for more variation in student achievement than any other factor (Darling-Hammond & Ball, 1997); therefore, the study of the teacher evaluation process is an important tool for supervisors. If researchers can pinpoint strengths and weaknesses in the process of teacher evaluation, then administrators can use that information more effectively to enhance a teacher's skills and, in turn, empower that teacher to increase student achievement.

### *Problems of Evaluation*

Research has shown that teachers view summative teacher evaluation practices as a meaningless exercise as currently practiced in the majority of school districts (Danielson & McGreal, 2000; Duke & Stiggins, 1986). Specific problems with summative evaluation include limited administrator experience with the classroom, minimal and/or incomplete observations, teacher preparation of carefully rehearsed lessons for announced observations, and lack of a common vision among administrators in defining good teaching. Researchers contend that most evaluation systems are created for teachers without teacher input, do not reflect the research on best practices, and are carried out with no relation to the professional development of teachers (Danielson, 1996). Additionally, Duke and Stiggins (1990) believe that lack of teacher input is one of

the main reasons that teachers do not choose to support summative evaluation methods. The evaluation systems widely used today were developed in the 1970's and fall short of current state standard requirements and learning research conducted in the past decade (Peterson, 1982). The current system relies upon the building principal to make determinations about all practices and allows little room for teachers to participate in their own evaluation (Danielson, 2000). Many evaluation procedures are based on the Madeleine Hunter Model (Danielson, 2000). Supporting research for this model involved low-level recall, and norm-referenced information. The Madeline Hunter Model does not stress the importance of current brain and learning research that maintains knowledge of achievement has evolved, therefore, so should evaluation of how teachers teach (Danielson, 2000). Another problem, according to Danielson, is that current evaluation systems depend on a single scale, such as satisfactory, needs improvement: 1, 2, 3; and low, medium, or high. These scales used in current evaluation systems do not give clear guidelines or rubrics clarifying what a "3" or a "satisfactory" rating constitutes. No differentiation is made between novice and experienced teachers. New teachers are held to the same standard as those who have had years to hone their craft. Most other professions build upon a system of apprenticeship. A surgeon does not become a surgeon overnight. A medical student spends many hours as an internist before becoming a surgical resident and witnesses or assists in dozens of procedures before earning the responsibilities and title of surgeon. Moreover, a recent vocational-technical graduate would not be held to the same standard as a master plumber. Why, then, should a first- or second-year teacher be evaluated on the same scale as a tenured veteran? Teacher assessment is used often to weed out the poorest performing teachers rather than improve

the performance of all teachers (Darling-Hammond, Wise, & Klein, 1999; Haney, Madaus, & Kreitzer, 1987).

The evaluation of teaching performance has long been regarded as an important vehicle for promoting student achievement. As one researcher commented, “The public has come to believe that the key to education improvement lies as much in upgrading the quality of teachers as in revamping school programs and curricula” (Darling-Hammond, 1990).

Scholars in the field have established that most teacher evaluation systems do not assist teachers to grow professionally, do not set high standards for teachers, and do not meet the needs of teachers trained in past generations (McGreal, 1983; Sergiovanni & Starrat, 1988; Shinkfield & Stufflebeam, 1995). Educators agree that professional evaluation should be a tool of continual pedagogical improvement for most teachers (Atkins, 1996). However, much of what has been written and presented about teacher evaluation reflects a negative perspective, as many teachers express anxiety and a feeling of uselessness brought about by the evaluation. Many teaching professionals dispute the relative importance of the two main purposes of evaluation: quality assurance and professional growth. Some argue that the two purposes are not compatible with one another (Danielson, 2000). The evaluation process at many school districts appears to be a simple bureaucratic requirement that yields little assistance to teachers and furnishes little information upon which administrators can base meaningful personnel decisions. Nevertheless, in other districts, the process of teacher evaluation is continually evolving, and teachers are taking part in their own evaluations. Although the most common forms of assessing teacher performance, such as summative and formative evaluation have

serious shortcomings, these forms are firmly entrenched in American schools. (McGreal, 1994; Medley, Coker & Soar, 1984). Coursework trains most educational administrators that teacher evaluation remains an integral part of their job duties. However, it is often difficult for administrators to dedicate enough time to generate and maintain a legitimate evaluative process in the face of the hectic daily schedule. Again, teacher evaluation is a key aspect of an administrator's obligation; nevertheless, it is not always a top priority given the many and varied challenges an administrator tackles. In trying to manage their many responsibilities, administrators often tend to focus on blatant unsatisfactory teaching practices. Because teacher evaluation is complex, sometimes viewed as threatening, and not very well understood by both teachers and administrators, much of the current practice involves shortcuts rather than systemic evaluation (Harris, 1986).

### *Teacher Perceptions*

Teacher perceptions of their evaluations are a driving force to school improvement, specifically in the area of school achievement and instructional effectiveness. As was previously stated in the empirical research on student evaluation and student achievement, teachers have a strong impact on instructional effectiveness. With this in mind, teachers' perceptions of their evaluations remain an important consideration when linking evaluation to achievement. If administrators can identify teacher perceptions on evaluation through research, those perceptions can serve as a catalyst to create connections in practice to school improvement, professional development, and student learning.

Bandura (1997) was the first to argue that one powerful construct varying greatly among schools and systematically related with student achievement is the efficacy of

teachers within the school. The relationship between teacher attitudes and teacher behaviors is critical to educational outcomes. Schools are social organizations made up of teachers who impact the achievement of students in their building. A faculty's collective sense of efficacy that they can promote high levels of academic progress contributes significantly to their school's level of academic achievement (Bandura, 1993). Schools where teachers have a high level of collective efficacy operate on the belief that students are "teachable" and can be motivated to achieve at high levels on national or state tests of language and mathematical competence. Efficacy has to do with perceptions of competence rather than actual level of competence. This is an important distinction because teacher efficacy is a significant predictor of productive teaching practices measured by teacher evaluation (Woolfolk & Hoy, 1990).

Empirical research on attitudes in the workplace indicates a strong and consistent link between self-efficacy and subsequent outcomes. One of the most important areas of study in organizational behavior is job attitudes, or how people feel about what they do when they work (Salancik, 1977). In education, too, employee acceptance of a performance appraisal or evaluation is crucial. The positive influences of attitudes of self-efficacy and job performance are well documented not only in education but also in other professions. For example, researchers have linked work attitudes toward evaluation to an increased level of task performance. Barling and Beattie (1983) noted that self-efficacy perceptions were strongly correlated to sales performance among life insurance agents. Subsequent studies have shown that attitudes can predict motivation and task performance in organizational settings (Ellis & Taylor, 1983). Self-efficacy is positively and strongly related to job performance (Bandura, 1986). Without such legitimacy,

supervisors are reluctant users who undermine the evaluation process (Lawler, 1969). Attitudes toward evaluation and the evaluator are important, particularly because the individual providing the performance appraisal is often the employee's supervisor (Bosswell & Boudreau, 2000). Furthermore, identifying key efficacy perceptions may be useful in determining which individuals will be high performers, specifically when little history of past performance exists or when past performance has been marginal (Gist, 1987).

Teachers perceive new systems of evaluation as being beneficial in their own continuous improvement. Turpin (2005) explored teacher perceptions of the teacher evaluation process in the Habersham County School District. The purpose of the study was to explore the teachers' perceptions of their new system of evaluation and determine if it was appropriately designed to increase the instructional effectiveness in the classrooms. Findings were that the new system of evaluation was perceived by the teachers to be beneficial in promoting continuous improvement. The teachers also expressed that they felt empowered by their evaluations, leading to the perception that they had a profound impact on student achievement.

In a 1994 study conducted by the National Center for Educational Statistics, eighty-nine percent of teachers believed that their last performance evaluation provided an accurate assessment of their teaching performance, and seventy-four percent thought that their last evaluation had been useful for improving classroom instruction.

Locally developed evaluation systems have been found to have a stronger impact on school improvement, professional development, and district reform initiatives. Colby (2002) studied the perceptions of teacher evaluation systems and their impact on school

improvement, professional development, and student learning. The study used twenty-one public schools in North Carolina. The sample size was 7,419 teachers. Completion of this study helped the school district identify areas where teachers needed professional development and contributed to attempts to increase student achievement.

To argue further that teacher perceptions of teacher evaluation should be researched, Argotsinger (2002) conducted an investigation of the performance-based teacher evaluation model and its impact on facilitating teacher growth in a school in suburban Missouri. That research, again, yielded the notion that certain practices as perceived by teachers facilitated teacher growth. Among the practices that teachers perceived as important were professional dialogue, collaboration, self-reflection, feedback, risk taking, and goal-oriented development plans. This study is significant because empirical research on teacher evaluation has demonstrated relationships between teacher performance on evaluations and student achievement. In addition, teacher evaluation has the potential to increase student achievement by encouraging the teacher to use effective teaching strategies, a significant implication for school districts in this time of educational reform and accountability.

Through a combination of formative evaluation and an assessment of teacher needs providing options for professional growth, teachers may be motivated to improve their practices positively influencing student achievement. Fowler (2001) researched teacher experiences with a new teacher evaluation program, one that included elements of teacher perceptions related to elements of teacher growth. Through data analysis, Fowler identified critical elements that influence teacher's perceptions of teacher evaluation practices. Three themes emerged with seven elements reported as either opportunities or

obstacles to teacher's perceived growth. The themes were professional learning culture, teacher autonomy, and program design. Teachers stated that these themes, when present, enhanced their knowledge of educational theory, as well as best practices, and applied learning to classroom instruction.

### Summary

Research demonstrates that teacher evaluation has been prevalent since the colonial times; however, evaluation systems differ and lack consistency. With NCLB, the question of accountability has come to the forefront: no longer can teacher evaluation be viewed as a bureaucratic requirement. Schools have been put on notice, and administrators and districts scramble to make changes and improvements to their current systems of teacher evaluation. Teachers are in a position to offer valuable information about how they perceive teacher evaluation. The information that teachers provide about their evaluations can assist administrators in decision-making, instructional preparation, and, ultimately, improvement in student achievement. If administrators can identify teacher perceptions on evaluation through research, these perceptions can serve as a catalyst to create connections in practice to student learning.

### *Purpose of the Study*

The purpose of the study was to understand the relationship between teachers' attitudes toward teacher evaluation in high- and low-achieving middle schools based on the Pennsylvania State System of Assessment. Several educational researchers (Darling-Hammond, 1986; Robinson, 1983; Weber, 1987; Frels, Cooper & Reagan, 1986) have shown evidence that the effectiveness of teacher evaluation can be significantly enhanced when teachers and principals view similarly the aspects of evaluation. Teacher

evaluations conducted in a positive environment hold the greatest potential to benefit both teacher and school district.

Teachers' attitudes toward teacher evaluation is an important research topic with implications for school districts reviewing their teacher evaluation systems. Teacher evaluation is extremely important in any efforts towards student achievement whether mandated or not by state or federal authorities. The research on effective schools identifies the importance of principals' instructional leadership behaviors in promoting higher levels of student achievement (Andrews & Soder, 1987; Edmonds, 1979). Leithwood and Steinbach (1994) defined instructional leadership as behaviors that directly affect supervision. Although teacher evaluation has gained attention recently, it remains a neglected area of educational research and practice even though evaluation practices have been labeled "seriously deficient" (Haefele, 1993), "chaotic" (Medley, Coker and Soar, 1984), and a "disgrace" (Scriven, 1994). The focus of this dissertation was to explore teacher attitudes toward teacher evaluation in high and low achieving middle schools in several districts. High and low achieving schools were identified by their Pennsylvania System of School Assessment (PSSA) scores and listing as a school meeting Adequate Yearly Progress (AYP) or not meeting AYP, as defined by NCLB. AYP is a part of NCLB, requiring accountability of schools to students, parents, teachers and the community. The purpose of AYP is to ensure that all students have reading and math skills that prepare them for the future. In addition, NCLB also states that all students must reach a proficient level or higher in reading and mathematics by the year 2014. School districts must show AYP on several measurable indicators such as attendance or graduation rate, test performance, and test participation. AYP measures

whether a school is making sufficient annual progress toward the goal of 100% proficiency.

The following questions were explored:

1. What are the attitudes of teachers toward formal evaluations?
2. Do attitudes of the teachers on teacher evaluation differ depending on the designation of a school's Adequate Yearly Progress status in the district in which they work?
3. Are there any significant differences in teacher perceptions of teacher evaluation based on sex, age, years of teaching, degree level?
4. Do teachers view the evaluation process as beneficial to teacher professional growth?
5. Do teachers view the evaluation process as one that is relevant to improve PSSA scores?
6. What are the major concerns that teachers have with the current system of teacher evaluation?

#### Definition of Terms

To assist the reader of this dissertation, essential terms are defined.

#### Evaluation

(1) A process that systematically examines the strengths and weaknesses of a teacher so as to provide benefit to both the individual and the organization (Casterter, 1981). (2) Any means that is used by school district personnel to assess the employee in terms that define the district or state education or classroom teacher requirements. (3) A process or procedure where administrators work with the teacher to determine his/her performance.

(4) According to Goldhammer (1980), evaluation involves a pre-conference prior to the observation to establish the targeted objectives and define what the teacher plans to accomplish in the lesson. Second, the evaluator actually observes classroom teaching for a predetermined period. Lastly, a post-conference examines the focus and other essential components of the lesson that may or may not have been brought up at the pre-conference.

#### Formative Evaluation

An evaluation method based on formal meetings between the evaluator and the teacher to determine the teacher's performance. Formative evaluation is designed to assist the employee in professional growth and to improve instruction. It is an on-going process of interaction between the evaluator and the teacher. The process identifies strengths and weaknesses of a teacher with the goal of improving the quality of teaching through the evaluation process (Millman, 1981; Stiggins & Bridgeford, 1984).

#### Summative Evaluation

(1) An evaluation method based on the summation of all methods used in the assessment of a teacher's performance. An evaluation may be summative and not formative, though both may occur in the process. (2) A process of collecting data to be used for decision making about the teacher's status within the organization. It provides information for administrative decision-making when dealing with hiring and retention, promotion and tenure, and salary matters (Millman, 1981; Stiggins & Bridgeford, 1984).

#### Attitude/Perception

(1) The understanding or misunderstanding that both principals and teachers have relative to the various aspects of teacher evaluation. (2) "The value judgments placed on the

behavior of people and/or activities according to the standards of the perceiver”

(Burchfield, 1986, p. 16).

### Observation

A method of collecting data about the teacher’s performance. The manner, forms, and frequency can differ depending on state and district guidelines. Observation as a method for evaluating teaching can take many forms. A broad range of instruments including behavior checklists, category systems, narrative records, summaries, and rating systems can be done (Millman & Darling-Hammond, 1990). Teacher observations are usually done in a classroom setting, but can be done in other venues such as cafeteria, recess, staff meetings, and conferences with parents.

### Pennsylvania Accountability System

System of assessment of all public schools and districts enacted in 2002 to comply with NCLB. This system is based upon content and achievement standards developed over a period. This term is another word for test. Beginning in the year 2002-2003, schools must administer tests in each of three grade spans: 3 to 5, 6 to 9, and 10 to 12 in all public districts. Beginning in the 2005/6 school year, tests were required to be administered every year in grades 3 through 8 in math and reading. In 2007, science will be tested. Along with achievement standards, school and districts are measured on performance such as attendance and graduation rates. The Pennsylvania Accountability System meets the requirements of NCLB. Schools are evaluated on a minimum target level of improvement—AYP. This system allows for schools to be measured in two ways; school’s absolute achievement (proportions of students who score at or above the proficient level) and a school’s growth in achievement from one year to the next to be

recognized. Achievement targets in reading and mathematics, as well as graduation, attendance and test participation. NCLB requires states to determine annually whether school districts and schools make AYP. Adequate Yearly Progress (AYP) is part of the federal No Child Left Behind Act and makes districts and schools accountable to students, parents and educators. To meet AYP, schools must meet achievement targets in reading and math, as well as graduation, attendance and test participation targets (See Appendices C & D.) (*Pennsylvania Accountability System*).

#### Value-Added Assessment

A new way of analyzing test data to measure teaching and learning. Based on a review of student test score gains from previous grades, researchers can predict the amount of growth students are likely to make in a given year. Thus, value-added assessment can show whether particular students—those taking a certain Algebra class, for example—have made the expected amount of progress, have made less progress than expected, or have been stretched beyond what they could reasonably be expected to achieve. Using the same methods, one can look back over several years to measure the long-term impact that a particular teacher or school had on student achievement (*Pennsylvania Accountability System*).

## CHAPTER TWO: REVIEW OF THE LITERATURE

### History of Evaluation

Due to the landmark No Child Left Behind Act of 2001 (NCLB), the practice of teacher evaluation is evolving further because it forces an analysis of current practices. A demand for higher levels of student achievement has left schools and their faculties under public scrutiny. Currently, teachers and administrators are being held accountable for lack of progress on state tests. Districts have implemented initiatives such as the 4 Sight Benchmark Assessment to measure how students will perform on the Pennsylvania System of School Assessment (PSSA). The tests predict how students will perform and provide useful data for focusing professional development and instruction (*Success for All Foundation*). The Success for All Foundation is a nonprofit organization that focuses on at-risk children. The goal of the organization is to create and implement programs based on research. The assessments used have the exact format, and cover the same information as the state assessments. The overall scores that these tests yield predict how students will perform on the PSSA (*Success for All Foundation*).

NCLB presented sweeping changes to education reform. Designed to improve student achievement and close achievement gaps, NCLB encompassed several ideals such as accountability, scientific research, parental options, and expanded local control and flexibility (U.S. Department of Education, 2004).

NCLB requires that every student across the country achieve his or her grade level in reading and math by 2014. Regardless of their race, ethnicity, or socioeconomic status, students must be tested in reading and math each year in grades three through

eight, and once in grades ten through twelve. The testing results must be shared with the district prior to the start of the following school year so that adaptations may be made to instruction. Data shared with the district must include disaggregate information, not only by school, but also by race, socioeconomic status, and special education classification.

Additionally, NCLB has a requirement for the hiring of highly qualified individuals. Teachers are mandated to have the appropriate qualifications/certifications such as a Bachelor's degree and a solid background in the content area in which they teach. Although no specific requirements under NCLB pertain to teacher evaluation, several school districts focus outcomes of teacher evaluations using the guidelines of what NCLB designates as important to student learning. States such as Pennsylvania have published the names of schools that are on the state's needs improvement lists in the media. This has further added to the public disdain for education reform.

This chapter provides a review of the literature relevant to the topic of teacher evaluation. It examines teacher evaluation from a historical perspective to illustrate what improvements or barriers have existed in teacher evaluation throughout history. The purpose of teacher evaluation is explored, as well as the problems encountered when conducting teacher evaluations. The chapter also discusses the common methods of teacher evaluation to help the reader understand the varieties of evaluation implementation by principals.

Teacher evaluation has been in existence since the early colonial period, when community members required faculties to maintain responsibility for standards. Peterson (1982) stated that the primary purpose of early teacher evaluations was to terminate incompetent teachers and ensure conformity to community standards. Peterson added

that evaluations dealt more with local societal beliefs, shared religious principles, or doctrines than with defined curricula. Primarily, the purpose for teacher evaluation provided a rationale to dismiss weak teachers (Peterson, 1982). According to Clark (1993), teacher evaluation grew out of a community's need to determine the future employment of its teachers, to assist in resolving compensation increases, and to influence promotion.

In the book, *A Century's Growth in Teacher Evaluation*, Peterson (1982) reported that from 1877 – 1900, men viewed as having knowledge rather than experience in the teaching field operated in the role of evaluator. For the most part during this era, non-academic school personnel subjectively performed teacher evaluation. The middle 19<sup>th</sup> century saw evaluation progress due to the growth of the population and to changes in the schools. This growth required more supervision by professionals, not lay people. As teaching evolved into more of a profession, evaluation rose to a new standard.

The Industrial Revolution, transforming American society, also gave teachers more power to decide their place in education. As farms gave way economically to small industry and the population continued to grow, large highly populated schools replaced small local schools. This change engendered a greater need for qualified teachers and supervisors to evaluate them. Not only did the Industrial Revolution cause the country to grow, it also expanded the existing knowledge base. The burgeoning knowledge base expanded the curriculum in schools, thus expanding subject-area expertise requirements, creating the need for educational leadership and teacher supervision.

In the mid-19<sup>th</sup> century, localities hired superintendents to oversee teacher supervision. Soon, the enormous responsibilities of superintendents precluded their

capacity to supervise all teachers; thus the role of the principal came to fruition. Principals gained the responsibility for the daily operation of the schools (Ellsbree, 1939). According to Chamberlain (1913), school reform came from reorganization of national and state governments. State officers were given power to dictate over schools, and this created the need for a public officer to oversee the schools. During the late 19<sup>th</sup> century, supervision of teachers was informal, unstructured, and not standardized (Bolin & Panaritis, 1992). This developed into a problem, because at the time, the population continued to grow and the laborers began to lack the skills necessary to compete in the new industrialized community (Karier, 1982). Secondary education grew to be necessary for all, not just the wealthy.

Around the turn of the 20<sup>th</sup> century, supervision emerged as a field of practice in response to increased levels of bureaucracy in school and the public demand for more control over the curricula (Karier, 1982; Bolin, 1987). Mostly informal, this early supervision failed to hold the teacher accountable to any great degree. According to Elsbree (1939) in *The American Teacher: Evolution of a Profession in a Democracy*, ministers or prominent citizens inspected schools. The “inspection,” as it was called, included the following purposes:

1. To appraise the progress made by children mastering the principles and doctrines of the Christian religion.
2. To hear the younger scholars read.
3. To appraise the performance of the scholars in writing and arithmetic.
4. To examine the writing of the masters.
5. To report on the number of scholars in attendance (p. 11).

With the beginning of the 20<sup>th</sup> century, people tried to apply the scientific approach to teacher evaluation (Peterson, 1982). This model came from industry and emphasized information, skills, habits, growth in attitude, etc. (Peterson, 1982). Emerging theories of personality in psychology influenced teacher evaluation. Researchers focused on the personality characteristics of good teachers (Ellett & Garland, 1987). During this time, an increased interest in educational research centered on teacher-training programs (Charters & Waples, 1929). During the 1920's, research on teacher evaluation began to measure factors such as student achievement, personality, intelligence, experience of the teacher, and training (Peterson, 1987). In the 1940's, World War II sparked an interest in establishing a more purposeful approach to teacher evaluation. The work of Dwight-Everett Beecher (1949) focused on salary and post-war inflation. Specifically, Beecher addressed the concerns about a fixed salary schedule for teachers that increased each year without respect to effectiveness. At this point in public education, most found this concept hard to accept.

In the 1960's, the deepening Cold War shifted the nation's focus to education and teacher evaluation due to the fact that the United States federal government determined to win the competition with the Soviet Union to gain power in the areas of science and research (Darling-Hammond, Wise & Pease, 1983). Peterson (1982) also states that the imperative for teacher evaluation originated from this Cold War philosophy and from discoveries in the field of psychology. This period saw the advent of collective bargaining, and of teachers' attempts to protect themselves from unfair evaluations (Peterson, 1982). According to Strike and Bull (1981), faculty union contracts affected the evaluation process by specifying the frequency of evaluations, informing teachers of

evaluative criteria, restricting methods for data collection, and specifying who could and could not participate in the evaluation process. A variety of instruments, such as the behavior checklist, category systems, and rating scales, were also developed during this time (Danielson & McGreal, 2000).

The 1970's and 1980's led to research on the need for change in teaching styles. Changing views of teaching styles impacted the way teachers planned and taught lessons. The use of various models of evaluation, such as the Hunter Model, produced a basis for observation instruments measuring teacher behavior and teacher effectiveness (Stallings, 1978). Extensive studies on teacher evaluation began to emerge as the Consortium for Research on Educational Accountability and Teacher Evaluation was born. The consortium was an international group committed to improving the evaluation of educators and educational programs. During this period, questions arose about the validity of teacher exams and job related competency. States began to move toward the development of classroom-based observation instruments, most being useful as effective measures of teacher behaviors through research (Stronge, 1997).

From the 1990's until today, teacher evaluations shifted to large-scale, statewide teacher evaluations addressing issues pertaining to classroom-based assessments (Ellett & Garland, 1997). During the 1990's, various areas of educational and cognitive science research, including critical thinking, content knowledge, content pedagogy, multiple intelligences, and cognitive learning theory, influenced teaching (Danielson & McGreal, 2000). In the 21<sup>st</sup> century, The Framework for Teaching Model came to fruition, based on the PRAXIS III: Classroom Performance Assessments criteria developed by the Educational Testing Service. This framework identifies those aspects of a teacher's

responsibilities documented through empirical studies and theoretical research as promoting improved student learning. These responsibilities seek to define what teachers should know and be able to do in the exercise of their profession. In this framework, the complex activity of teaching divides into 22 components clustered into four domains: planning and preparation, classroom environment, instruction, and professional responsibilities. This framework of professional practice is designed to meet the needs of novice teachers concerned with day-to-day survival, experienced teachers wanting to improve their effectiveness and help their colleagues do so as well, and highly accomplished teachers wanting to move toward advanced certification and serve as a resource to less-experienced colleagues. Evaluative criteria are provided for each domain and its components; instead of rating specific behaviors, these criteria refer to a set of commonalities that reflect or support actions (Danielson, 2000). Simply stated, beginning with the 1960's through today, accountability has been a major focus of teacher evaluation.

#### Purpose of Evaluation

The actual purposes for evaluating teachers are numerous and conflicting. They can range from mandates, school improvement, to inadequacy of teaching staff. Despite the fact that most of the literature revealed that ultimately the purpose for teacher evaluation should be accountability, Bolton (1973) stated that the purposes of teacher evaluation were as varied as the school's need to satisfy the public's demand for accountability to the individual teacher's desire to improve the way he teaches a particular aspect of one subject area.

Natriello defined teacher evaluation as one of many processes in schools used to legitimate the organization and its existence (ctd. in Millman & Darling-Hammond, 1990). He further stated that there were three major purposes of evaluation:

1. Influencing the performance of individuals within particular positions;
2. Guiding decisions about the movement of individuals into and out of positions, granting tenure or firing, movement to a new office or level; and
3. Legitimizing the organizational control system itself; that is, persuading employees and constituents of the organization that personnel activities are performed fairly and effectively (p. 35).

Duke and Stiggins (1986) defined teacher evaluation as a process providing information for use in personnel management decisions such as hiring, firing, promotion, tenure, salary, or merit pay. Teacher evaluation was generally thought of as an administrative function used to judge the effectiveness and quality of instruction delivered by an employee. Peterson (1995) added that “evaluation is intended to protect children, reassure teachers that they are doing good jobs, assure audiences interested in teacher performance, make personnel decisions, inform teacher educators, and shape further practice” (p. 30).

Scholars in the field of teacher evaluation listed many purposes for evaluation. The Personnel Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (1988) stated that nine purposes existed for evaluating teachers. These nine consisted of the following: entry to training, certification/licensing, selection, performance reviews, staff development counseling, merit awards, tenure decisions, promotion decisions, and termination. Similarly, the Educational Research Service

reported, *Teacher Evaluation: Practices and Procedures* (1988) listed the following as major purposes for teacher evaluation: to ensure that all teachers are at least minimally competent and to recognize the performance of outstanding teachers (p. 1). Many others including Stufflebeam (1992), Darling-Hammond (1990), and McLaughlin and Bernstein (1994) affirmed the above-mentioned items in their scholarly literature.

In 1996, a study was conducted in 42 public schools in southwestern Pennsylvania surrounding the city of Pittsburgh (Tomaino, 1997). The researcher concluded that teacher perceptions of the purpose of teacher evaluation is different than those of other stakeholders, such as administrators or board members. In his study, Tomaino found statistically significant differences among the perceptions of teachers, principals, superintendents, and board members relating to accountability practices as being one of the primary purposes of teacher evaluation. This falls in line with other authors' arguments that misunderstandings persist between teachers and administrators on the purposes of teacher evaluation (Duke, 1987).

Several researchers (Gibb, 1989; McGreal, 1982) have argued that the accountability aspect of teacher evaluation be de-emphasized. Furthermore, McGreal (1982) cautioned that systems of evaluation based strongly on accountability promoted negative feelings, created anxiety, and lowered the likelihood of teachers cooperating and adjusting their classroom instruction. Natriello (1990) went as far as to say that teacher evaluation systems were one of many control processes in schools used to legitimize the organization and its existence. That author described the three following purposes of teacher evaluation as related to "power" and "control":

1. To control or influence the performance of individuals within particular positions
2. To control movement in and out of positions, and
3. To legitimize the organizational control system itself (p. 36-37).

#### Problems of Evaluation

The two most frequently cited purposes of teacher evaluations were accountability and performance improvement (Duke, 1990). Teachers realize that evaluations are necessary and required in the Commonwealth of Pennsylvania; however, teachers have some reservations about the process and its use. As Duke and Stiggins (1988) indicated, states have re-assessed their evaluation policies and practices. This, again, works to confuse teachers as to the true purpose for teacher evaluation. Fullan (1996) suggested that one reason for systemic failure of reforms like teacher evaluation is fragmentation: “Fragmentation occurs when the pressures--and even the opportunities--for reform work at cross purposes or seem disjointed and incoherent” (p.420). This concerns not only teachers, but also principals, because a conflict inherently exists in the principal’s role in conducting the evaluation.

Although research has indicated new findings about teaching and learning, teacher evaluation practices have been stagnant (Ellett & Garland, 1987; Holley, 1979). Ellett and Garland found that:

1. Teacher evaluation had not changed since 1979.
2. Administrators were ignorant of recent developments in teacher evaluation.
3. Large districts stressed summative uses of data more frequently than formative uses.

4. Few districts in this 100-district study reported practices that included performance standards or orientation for evaluators.
5. Local evaluator training was insufficient.
6. Local teacher evaluation instruments included criteria that were in opposition to the research literature (p. 75).

Research studies (Loup, Garland, Ellett, & Rugatt, 1996; Medley et al., 1984) indicated that principals and other administrators sometimes increase their evaluation scores to prevent jeopardizing the relationship that they have with the teacher. The principal must wear two hats and tread lightly when evaluating teachers. In one regard, the principal has to be trusted and seen as a person who supports teachers and their professional growth; on the other hand, because of high stakes testing, the principal may also have to put teachers on improvement plans or terminate them if they do not produce the desired student achievement results. This presents a dilemma for all parties involved.

Most assessment tools used in the past and present were based on the notion of generic teacher characteristics. In addition, classroom observation instruments are generic since they are used across all grade levels and subject areas. This is problematic because teacher behaviors vary greatly across content areas, grade levels, and subject. For instance, some districts use classroom observation records to assess the school nurse, psychologist, and guidance counselor who may not be observed in a classroom setting. The specific knowledge required in different contexts is so different that the same assessment tool cannot capture it (Madus & Pullin, 1987; Haertel, 1991).

In 2000, Raymond A. Bastarache conducted a research study in Massachusetts at three urban school districts to determine to what extent teachers and principals agreed on

the purposes, methods, and effectiveness of teacher evaluation. The researcher found that principals and teachers do not agree that teacher evaluations improve teacher skills in the area of instructional techniques. The teachers also felt that their evaluations had little or no impact on improving their classroom performance in the classroom.

Similarly, Stiggins and Bridgeford (1984) stated that a brief, yearly observation performed by unequipped administrators had little or no impact on teacher or school improvement. Frase and Streshly (1994) also believed that central office administrators seldom used evaluation data from school to make necessary or meaningful improvements in classroom instruction. Finally, many authors believed that observations done in classrooms are not necessarily representative of daily classroom activities (Darling-Hammond, Wise, & Klein, 1999). Often, the teacher is aware of the observation date; therefore, the he or she can prepare a canned lesson: one that he or she knows will meet the administrator's evaluative criteria. Also, collective bargaining agreements may prohibit administrators from implementing necessary frequent observations to assess teacher performance. In some districts, union contracts make it clear when administrators can and cannot observe and exactly what they can write in reports. Consequently, teaching fails to be adequately monitored or evaluated (Tucker & Stronge, 2005). As Schmoker (2006) stated, "The tradition of evaluating teachers on the basis of one or two observations is a farce" (p. 137).

Natriello (1984) conducted research pointing to the belief that many teachers viewed evaluations as a way for administration to maintain control over classroom instruction and dictate what and how to teach their subject area. Adding to this, Newton and Braithwaite (1988) suggested that teachers perceived evaluations as a means of

bureaucratic control by administrators. Confusion about teacher evaluation usually occurred when opinions differ concerning the purpose of the evaluation (Battaglia, 2001). In some ways, administrators may view evaluation as a means to manage and control staff or as a means of accountability to his or her respective boards. Teachers, conversely, may view evaluation as a means to engender professional growth. Then again, community members and parents may expect that teacher evaluations should guarantee “the appropriate treatment of children in classrooms” (Mitchell, Wise, & Plake, 1990). How interested parties perceive evaluations must be reviewed in order to improve the usefulness of teacher evaluations.

Researchers concluded that administrative reports of classroom performance were inaccurate because of inadequate reliability and validity (Medley, Coker, & Soar, 1984; Scriven, 1994; Stodolsky, 1984). In addition, other tasks required by the school system often constrained principals who normally carry out observations. Scriven (1981) described the problems of accuracy associated with evaluations. He included six factors in his description:

1. Change in the usual teaching practice caused by the visit itself.
2. Unreliable number of samples.
3. Personal biases of the evaluator.
4. Adult raters who do not think like students.
5. Style preferences of the evaluator.
6. Costs in time of lengthy classroom visits (pp. 244-271).

Scriven also noted that the evaluator rarely reviews curriculum content, student achievement data, or professional collaboration.

Wood and Pohland (1979) found that rating forms used in teacher evaluation can discourage accurate, valid reporting of teacher performance. In Wood and Pohland's study of 65 teacher-rating scales, the teaching role of teachers (i.e., adequate preparation of lesson plans, evaluation of student progress, higher-level challenges) constituted only 28% of the total items. The majority of scale items evaluated the control of teacher behavior, teachers' personal characteristics, and teacher involvement in non-instructional roles such as volunteering for school committees.

Teacher evaluation systems also lacked variable data sources (Peterson, 1984). This lack is important because teachers are good teachers for different reasons. No single data source is valid for every teacher in a school, and no individual data source is available for all teachers (Travers, 1981). Peterson (1995) also strongly criticizes standard teacher evaluation stating that:

Seventy years of empirical research on teacher evaluation shows that current practices do not improve teachers or accurately tell what happens in classrooms. Administrator reports do not increase good teachers' confidence or reassure the public about teacher quality. Teacher evaluation as presently practiced does not identify innovative teaching so that it can be adopted by other teachers and used in teacher education programs. Finally, current procedures do not reward exemplary teachers (p. 14).

Other research findings on evaluation related the lack of evaluator competence and of time for adequate observation/feedback; teacher and administrator understanding and acceptance; emphasis on minimum competency; political viability; and the application of narrow conceptions of teaching (Peterson, 1995). Additionally, Peterson cited several

studies illustrating the statistical inaccuracy of principal ratings and low levels of respect for evaluation procedures among administrators and teachers. Wise, Darling-Hammond, McLaughlin, and Bernstein (1984) found that principals lacked sufficient resolve and competence to evaluate accurately, primarily stemming from role conflict between their responsibility as instructional leader and as staff supervisor.

Finally, in a study of principal ratings of teacher performance, Nelson & Sassi (2000) found that principals tended not to delve deeply into issues of content-specific pedagogy. Instead, principals focused classroom observations on more generic teaching processes. The authors suggested that, in order to assess teacher performance accurately on standards-based instruction, evaluators must understand content matter thinking in which students engage and methods by which teachers can foster content knowledge skills.

### Common Methods of Teacher Evaluation

Research has not identified a teacher evaluation method successful for every teacher. Darling-Hammond, Wise & Pease, (1983), Shulman, (1988), and Darling-Hammond (1995) suggested that taking a single-dimension approach toward teacher evaluation did not produce sufficient information on a teacher's performance, competence, or effectiveness. Historically and currently, the direct observation of teachers with checklists and rating scales remained the mainstay of the evaluation process (Darling-Hammond, Wise & Pease, 1983).

By order of state law and/or regulation, almost all American teacher evaluation systems outlined a formal procedure for the evaluation of teachers (Danieleson & McGreal, 2000). Most teacher evaluation systems attempted to combine the two

functions of formative and summative evaluation. These very broad evaluations provided a database for administrative decision-making such as dismissal, promotion, and tenure (Stiggins & Bridgeford, 1984). Popham (1986) referred to these methods as a “dysfunctional marriage” and believed they were counterproductive to the intended purpose. Major criticisms of these two evaluation processes included the fact that the same individual frequently conducted them both at the same time (Popham, 1986; Christen & Murphey, 1987).

The goal of formative evaluation was to develop and improve teacher performance. According to Duke and Stiggins (1986) the principle of formative evaluation was based on the teacher’s ability to make improvements in his or her knowledge and skills. Examples of formative teacher evaluation models included clinical supervision, peer supervision, portfolios, and artifacts. Three features defined formative evaluation: it was on going; it involved formative assessment; and it sought in-depth information about a teacher’s practices (Beyer, 1995). Beyer also stated that formative evaluation served developmental purposes, with the primary goal to shape and improve the quality of a person’s skills. Some problems occurred with the use of formative evaluations. Beyer (1995) identified six flaws apparent in many formative evaluations:

1. Using unqualified people as expert evaluators,
2. Failing to give specific or appropriate guidelines to evaluators on what to look for,
3. Asking for general opinions or judgments without defining criteria,
4. Failing to see the kinds of information most useful in improving the quality of teaching,

5. Failing to continuously assess an educator throughout the course of development, and
6. Failing to allow enough time for the evaluation and the use of its results.

Summative teacher evaluation was best defined as a model to assist in decision making such as hiring, promotion, and termination of teachers (Scriven, 1981). Most teacher evaluation systems were summative in nature and promoted educational accountability by judging teacher effectiveness (Stiggins & Bridgeford, 1984). Summative evaluation is a process of collecting data used for making decisions about a person's future status within an organization (Danieleson & McGreal, 2000). As previously stated in Chapter One, this procedure served as the mainstay for the early teacher evaluation process and is still used in many public schools today. Some of the characteristics associated with summative evaluation included attitudes, dependability, responsibility, appropriate use of instructional materials, planning, modeling, and student evaluation. Summative evaluation centers on direct observations conducted by administrators a few times per year. Summative evaluation forces the evaluator to place a "value" judgment in the evaluation process.

However, research showed that teachers viewed summative teacher evaluation practices as a meaningless exercise (Danieleson & McGreal, 2000). Researchers also believed that this practice did not elicit institutional improvement (Wise, et al., 1984). Among teachers, one of the biggest concerns with summative evaluation revolved around the lack of common vision among administrators in defining good teaching (Danieleson, 1996). Peterson (1995) believed that evaluation systems were designed to assess minimal competency and omitted positive feedback on exceptional teaching skills. Teachers

complained that their principal was not knowledgeable about the characteristics of their subject matter or grade level (Duke et al., 1986).

The use of summative evaluation by administrators has been entrenched in our educational system from the beginning of formalized education. The literature over the past twenty years highlighted the downfalls of summative teacher evaluation; however, this practice currently persists in schools today (Darling-Hammond, 1990). Evaluations look the same district to district, and teachers and administrators go on widely and quietly accepting these poor practices in teacher evaluation (Peterson, 1995).

### Summary

The literature review has provided the rationale for the importance of teacher evaluation in the age of accountability. Various findings indicated the dissatisfaction and distrust held by teachers in the area of teacher evaluation. A consensus existed in the literature that teacher attitudes about evaluation differed tremendously. Research also indicated that purposes for teacher evaluation were numerous and varied and that confusion reigns not only for teachers but also for administrators. Two forms of evaluation stood in the forefront: formative and summative evaluation. Studies suggested problems with instruments used for evaluations, particularly instruments used across all subject areas. Researchers also concluded that reports on class performance were inadequate due to reliability and validity issues. Accountability in education, specifically to fulfill NCLB, requires teachers to think seriously about their evaluations and improvements to instruction; thus, this finding supports the purpose of this study. The literature review established many questions; nevertheless, researchers agreed that with high-accountability requirements of schools, and specifically of NCLB, current teacher

evaluation systems must be improved, and stakeholders must collaborate to accomplish this improvement.

## CHAPTER THREE: METHODOLOGY

### Purpose of the study

The purpose of the study was to understand the relationship between teachers' attitudes toward teacher evaluation in high- and low-achieving middle schools based on the Pennsylvania State System of Assessment. Several educational researchers (Darling-Hammond, 1986; Robinson, 1983; Weber, 1987; Frels, Cooper & Reagan, 1986) have shown evidence that the effectiveness of teacher evaluation can be significantly enhanced when teachers and principals view similarly the aspects of evaluation. This important research topic can have implications for districts not meeting Adequate Yearly Progress (AYP). Current school reform movements, such as NCLB, have put teacher effectiveness at the forefront of school improvement. As the literature review stated, teacher evaluation should figure prominently in any effort toward improving student achievement, whether mandated by state and/or federal authorities or not. Teachers' attitudes toward teacher evaluation should have implications for school districts reviewing their teacher evaluation systems. Effective schools research identifies the importance of teacher evaluation; however, it remains a neglected area of educational research and practice even though evaluation practices have been labeled "seriously deficient" (Haefele, 1993), "chaotic" (Medley, Coker & Soar, 1984), and a "disgrace" (Scriven, 1994).

As stated in the introduction, the relationship between teacher attitudes and teacher behaviors is critical to educational outcomes. Schools are social organizations

comprised of teachers influencing student achievement in their buildings. A faculty's collective sense of efficacy enables that faculty to promote high levels of academic progress contributing significantly to their school's academic achievement level (Bandura, 1993).

The definition of high- and low-achieving schools was based on the Pennsylvania State System of Assessment (PSSA), specifically middle schools that met or did not meet AYP as defined by NCLB. AYP, a measure of yearly progress toward achieving state academic standards, defines the minimum level of improvement that states, school districts, and individual schools must achieve each year.

#### Research Questions

Middle schools comprised this study because all students in the 8<sup>th</sup> grade were tested according to curricular standards in mathematics and reading. This study attempted to determine if a link existed between higher PSSA scores and teacher attitudes toward their teacher evaluations. Were the views of teachers regarding their evaluations directly related to schools meeting AYP or not meeting AYP? Can one assume that scores were a direct result of some other factor(s), such as demographics, location, and/or size of district? The specific research questions included the following:

1. What are the views of teachers toward formal evaluations?
2. Do attitudes of teachers toward teacher evaluation differ depending on the designation of a school's AYP status in the district in which they work?
3. Are there any significant differences in teacher perceptions of teacher evaluation based on sex, age, years of teaching, or degree level?

4. What are the views of teachers concerning the evaluation process and its contribution to teacher professional growth?
5. What are the views of teachers on the evaluation process and its relevance in improving PSSA scores?
6. What are the major concerns that teachers have with the current system of teacher evaluation?

Results from this study will provide data for school and school district decision-making regarding teacher evaluation processes. The findings will enhance understanding of the concerns about the process and promote mutual understanding, better planning, and development of teacher evaluation practices. Furthermore, one hopes that schools that do not meet AYP will have the ability to review the data and make the necessary changes to their evaluation processes.

#### *Research Hypotheses*

- H1: There are negative attitudes of teachers towards teacher evaluation.
- H2: There are significant differences in teacher attitudes towards teacher evaluation between schools that meet AYP and those that do not.
- H3: There are significant differences in teacher perceptions of teacher evaluation based on sex, age, and teaching experience.
- H4: Teachers view the evaluation process as beneficial to teacher professional growth.
- H5: Teachers view the evaluation process as relevant to improve PSSA test scores.

H6 There are no major concerns that teachers have with the current system of teacher evaluation.

### *Study Variables*

The dependent variable in this study was teacher attitude towards evaluation as measured by the Teacher Evaluation Profile (TEP). The independent variables included the following: (a) AYP measured as met or not met, (b) teacher's gender, (c) teacher's age, (d) teacher's years of experience, and (e) teacher's degree level. The independent variables were coded as follows:

- AYP (0 = not met, 1 = met)
- Gender (male = 1, female = 2)
- Age=Age
- Years of experience (0-5 = 1, 6-10 = 2, 11-15 = 3, 16-20 = 4, 21-25 = 5, 25+ = 6)
- Degree Level (Bachelor = 1, Master = 2, Doctorate = 3).

### *Population*

The study was limited to middle schools in southwestern Pennsylvania due to the fact that all students in the 8<sup>th</sup> grade were tested according to curricular standards in mathematics and reading. There are 68 middle schools in Allegheny County. Eighteen are in the City of Pittsburgh School District. Out of the eighteen in the City of Pittsburgh School District, five failed to meet AYP in 2006. Of the remaining 50 middle schools in Allegheny County, eight failed to meet AYP in 2006. Combining the city and the county, one finds that thirteen of 68 middle schools failed to make AYP. This study anticipated that 400 teachers would respond to the survey from AYP and non-AYP schools. The samples were stratified further by suburban, rural, and city schools.

### *Method of Data Collection*

The researcher gained permission to use the Teacher Evaluation Profile (TEP) from the original authors of the survey instrument. Then, the researcher used the Pennsylvania Department of Education's website and database for middle school AYP data for the 2005-2006 school year. This data is available on the Internet on the Pennsylvania Department of Education's website at [http://www.paayp.com/state\\_report.html#](http://www.paayp.com/state_report.html#). Data collection occurred during the spring of 2007. Teachers were chosen randomly from the personnel directory lists from districts chosen to participate in the study. These teachers participated voluntarily and were protected through confidentiality. The survey instrument was coded according to those middle schools that demonstrated AYP (white paper) and those middle schools that did not demonstrate AYP (blue paper). Along with the instrument, potential respondents in the sample were mailed a postage-paid envelope to utilize in returning the completed instrument. Population members were given two weeks from the date on the cover letter to complete the questionnaire. Of 400 surveys sent, 79 were returned. This accounted for a return rate of 20% which is consistent with other educational research. It was consistent with other studies conducted in the social sciences that obtained return rates in the low 20's (Porter & Whitcomb, 2003).

### *Instrument Development*

The survey instrument, a modified version of the TEP developed by Richard Stiggins and Daniel Duke at the Northwest Regional Educational Laboratory in Portland, Oregon, measured the experience of a teacher's evaluation. The TEP was designed to validate a list of key attributes of teacher evaluation by determining if these attributes

related to perceived growth outcomes of evaluation. The five attributes were (a) the teacher, including interpersonal manner and teaching experience; (b) the person who evaluated performance; (c) the information gathered on performance; (d) the feedback received; and (e) the evaluation context (Stiggins & Duke, 1988).

The survey was based on the outcomes of three sets of studies that identified 44 attributes of the teacher evaluation process. The first study commenced with a study of teacher evaluation systems in four school districts. The districts' purpose was to "uncover barriers to teacher growth through effective evaluation" (Stiggins & Nickel 1988). The barriers identified by teachers and administrators include lack of training among participants in effective evaluation and feedback procedure, insufficient time available for evaluation, lack of mutual trust between teachers and supervisors, and the accountability concerns associated with evaluation (Stiggins & Nickel 1988). The second study focused on teachers who reported that they had experienced professional growth because of their evaluation experience. The factors identified as adding to the evaluation experience are the teacher, the evaluator, the procedures, the feedback, and evaluation context (Stiggins & Nickel 1988).

The third study asked teachers to rate the overall quality and impact of their most recent evaluation; additionally, the survey asked the teachers to describe nine specific aspects of themselves as teachers. These nine aspects included their strengths and orientation to risk-taking. Finally, the teachers were directed to describe their perceptions of the person who evaluated their performance, in terms of that evaluator's credibility as a source of feedback on teaching, interpersonal manner, and knowledge of the technical

aspects of teaching. After the tests confirmed reliability and validity, the questionnaire was revised and refined to become the TEP (Stiggins & Nickel, 1988).

The TEP directed teachers to describe their most recent evaluation experience from a variety of perspectives. First, teachers answered a number of demographic questions related to their teaching situation. Second, teachers rated the overall quality of their evaluation using a 10-point scale from 0 to 9, with 0 representing very low quality and 9 representing very high quality. Teachers rated the overall impact of their evaluation experience on four educational practices using a 10-point scale from 0 to 9, with 0 representing very low impact and 9 representing very high impact. Last, teachers rated their perceptions of the existence and degree of use of the 44 attributes across four categories, each having a unique rating format based on a 1 to 5 scale, with 1 representing low on the particular scale and 5 representing high on the same scale. (see appendix b)

The TEP took approximately ten minutes to complete. Instead of using the original authors' rating scale of "A B C D E," this researcher used a rating format based on a 1 to 5 scale. This change was made to simplify the process for putting the results data into the statistical software program.

### *Data Analysis*

The researcher listed the results, percentage of responses, descriptive statistics, and appropriate comments to the responses and data analysis. Results were tabulated using SPSS statistical software. The nature of questions determined the statistical procedure employed. The following list includes some of the statistical procedures performed: descriptive statistics, independent t-test, analysis of variance and correlation analysis.

For research question #1, the mean scores for the survey items were obtained.

For research question #2, an independent t-test was used to compare those schools that met AYP and those that did not meet AYP. For research question #3, a combination of statistical analyses was employed. Correlation analysis was performed to find if perception was related to the age of the respondents. To compare if differences existed due to sex, an independent t-test was performed. To determine if differences existed due to years of teaching experience, a one-way ANOVA was completed. For research questions #4, #5, and #6, frequencies of the response categories were obtained.

## CHAPTER FOUR: RESULTS

The purpose of this study was to understand the relationship between teachers' attitudes toward teacher evaluation in high- and low-achieving middle schools based on the Pennsylvania State System of Assessment. Specifically, the study sought to answer the following six research questions:

1. What are the views of teachers toward formal evaluations?
2. Do attitudes of teachers toward teacher evaluation differ depending on the designation of a school's AYP status in the district in which they work?
3. Are there any significant differences in teacher perceptions of teacher evaluation based on sex, age, years of teaching, or degree level?
4. What are the views of teachers concerning the evaluation process and its contribution to teacher professional growth?
5. What are the views of teachers on the evaluation process and its relevance in improving PSSA scores?
6. What are the major concerns that teachers have with the current system of teacher evaluation?

To attempt to answer these questions, the TEP was administered to teachers in participating schools that met AYP and those that did not meet AYP. Of 400 surveys mailed out, 79 valid surveys were returned, giving a return rate of about 20%. Although this return rate is rather low, it is now common for educational researchers to use survey data in which the response rate is less than 50%. This practice is consistent with other

studies conducted in the social sciences that also obtained return rates in the low-20 % range (Porter & Whitcomb, 2003).

In the following section, results to individual research questions are presented.

*Research Question One*

What are the views of teachers toward formal evaluations?

Table 1: Descriptive Statistics for Teacher Perceptions of Evaluator

Teacher Perceptions of Evaluator	Mean	SD
Credibility	4.00	.93
Helper	4.00	.91
Trustworthy	4.00	.93
Threatening	4.14	.89
Impatient	3.86	.86
Rigid	3.82	.84
Knowledgeable	3.97	.88
Capacity	3.56	1.01
Familiarity	3.53	1.21
Experience	3.90	.96
Usefulness	3.56	1.11
Persuasiveness	3.52	1.06

To answer this question, descriptive statistics (means and bar graphs) were obtained for the items or group of items in the survey instrument. One of the ways in which the views of the teachers toward formal evaluation were measured on the instrument was through a battery of items measuring perceptions of teachers on the evaluator. Table 1 below presents means and standard deviations of items measuring teacher perceptions toward the evaluator. Perception was measured on a five-point Likert scale from 1 to 5, with 1 indicating low perception and 5 high perception. Table 1 shows that, overall, teachers had a positive perception on the evaluator. For example, nearly

four of five teachers think that their evaluator is credible, is a helper, is trustworthy, and is not threatening. It is important to note that the survey instrument asked questions not only about the person who evaluated them, but also, on information obtained during the evaluation and the feedback from the evaluation. There was nothing statistically significant to report in those other categories. In this particular study, it appeared that the evaluation process hinged on the evaluator.

Figure 1 represents the number of informal teacher observations per year. As seen in the bar graph, a majority of the teachers reported that they are informally observed less than one time per month. Only four teachers reported being informally observed on a daily basis; eleven, once per month; and four, once per week.

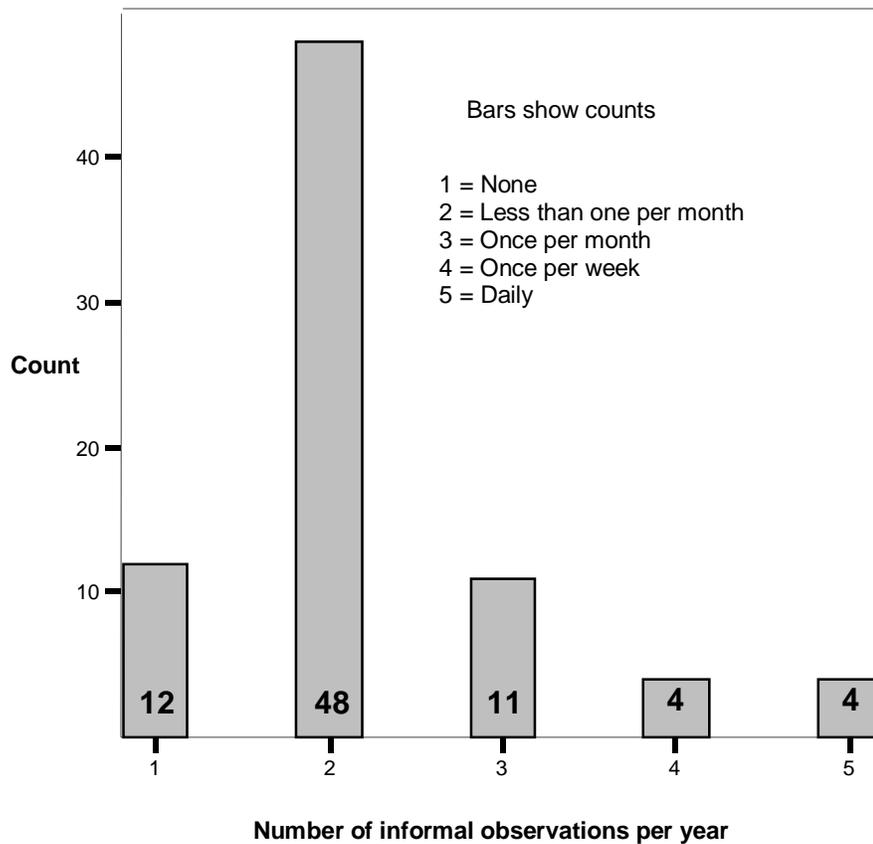


Figure 1: Number of Informal Teacher Observations

The figure represents information on the process of evaluation. According to these results, teachers were not observed on a regular basis. As the reader has seen in the review of the literature, teachers remarked that one of the problems with teacher perceptions of evaluation was the fact that they did not get observed on a regular basis.

*Research Question Two*

Do attitudes of teachers toward teacher evaluation differ depending on the designation of a school's AYP status in the district in which they work?

This question was answered by running an independent t-test. The attitudes differed on most of the items; however, they are only significantly different on three items in the table below.

Table 2: Independent T-Test Results on Teachers

	AYP (n=49) Mean	Non-AYP (n=25) Mean	t	p-value
Were standards endorsed	3.49	4.08	-2.249	.028
Exam of classroom records	2.96	3.68	-2.472	.016
Length of informal observation	2.57	1.48	4.022	.000

As one can see in the table, differences existed between AYP schools and non-AYP schools whether standards were endorsed or not. In non-AYP schools, the mean on the item 'Were standards endorsed' was 4.08 compared to 3.49 for AYP schools. This difference was statistically significant ( $t(72) = -2.25$ ,  $p = .028$ ).

In addition, differences existed between the two types of schools in the area of 'Examination of classroom records'. In non-AYP schools, the mean on the item 'Examination of classroom records' was 3.68 compared to 2.96 for AYP schools. This difference was statistically significant ( $t(72) = -2.47$ ,  $p = .016$ ). Non-AYP school teachers perceived that their evaluators spend more time examining classroom records than AYP school teachers.

Last, differences existed between the two types of schools on length of informal observation. In AYP schools, the mean on the item 'Length of informal observation' was

2.57 compared to 1.48 for non-AYP schools. This difference was statistically significant ( $t(72) = 4.022, p = .000$ ).

### *Research Question Three*

Are there any significant differences in teacher perceptions of teacher evaluation based on sex, age, years of teaching, or degree level?

To answer this question, several analyses were obtained. First, an independent t-test was performed to compare attitudes based on sex, the results of which are summarized in Table 3.

Table 3: Independent T-Test Results on Differences due to Gender

	Male n=31	Female n=48	t	p-value
Amount of time spent on evaluation process	2.71	3.19	-2.280	.025
Attributes of information gathered	1.86	2.43	-2.777	.007
Standards for evaluation	2.97	2.32	2.086	.040

Out of the six sections of the survey instrument, only three areas were statistically significant in relation to differences due to gender. The perception of female teachers in the study was that there was more time spent on the evaluation process than male teachers, and also that the attributes of information gathered on them was greater. On the other hand, male teachers believed that the standards for evaluation were considered important attributes for the information gathered.

Table 4: Correlation between Age and Perception

	Age	Perception
Age	1.00	.442**
Perception	.442**	1.00

\*\*Significant at the 0.01 level

To determine if any relationships exist between age and perception, a correlation analysis was obtained. Table 4 illustrates the analysis results. One can see that a

significant positive relationship exists between age and perception. As their ages increase, teachers tend to rate their competence higher. The correlation between perception and age is .44, which is positive.

To compare differences in perception based on years of teaching, an analysis of variance was performed. Table 5 displays the results of this analysis. All items in the table are significant for formal evaluation.  $F(2,78) = 9.76 ; p < .05$ . A positive relationship exists between years of teaching experience and perception. Those with zero to five years of teaching experience rate their teaching competency lower than those with more experience. Also, teachers with more years of teaching experience rate their evaluators as being more trustworthy than those teachers with fewer years of experience. After the ANOVA test was run and results were determined, a post-hoc Tukey test was run to determine where the differences lay. Most are between the least experienced group, 0-5 years and all other, but no differences are found among other groups.

Table 5: ANOVA Table for Differences in Perception due to Experience

	F-Ratio	p-value
Formal evaluation	9.76	.000
Informal evaluation	3.34	.014
Level of trust	2.81	.032
Interpersonal manner	2.94	.026
Flexibility	2.70	.037
Familiarity with classroom	2.97	.025
Frequency of formal feedback	3.37	.014
Frequency of informal feedback	3.79	.007
Depth of information provided	2.80	.032
Clarity regarding purpose of evaluation	3.04	.022
Impact of collective bargaining agreements on evaluation	2.52	.048

To compare differences in perception based on degree level, analysis of variance was performed, the results of which are shown in Table 6 below. Two items in Table 6 are significant in perceptions due to degree level. A post-hoc test was run to find out where the differences lay. The two significant differences are found in competence as a teacher  $F(2,78) = 5.96 ; p < .05$  and clarity of standards  $F(2,78) = 5.92 ; p < .05$ . The differences in perception due to degree level are significant between bachelor- and master-degree levels and between bachelor- and doctorate-degree levels. No significant differences existed between master-degree and doctorate-degree levels.

Table 6: ANOVA Table for Differences in Perception due to Degree Level

	F-Ratio	p-value
Competence as a teacher	5.96	.004
Orientation of others	4.04	.022
Knowledge of subject matter	4.51	.014
Credibility of feedback	3.28	.043
Working relationship	4.58	.013
Temperament of evaluator	4.18	.019
Familiarity with classroom	5.45	.006
Clarity of standards	5.92	.004
Information received	3.30	.042
Quality of ideas and suggestions	3.13	.049

*Research Question Four*

What are the views of teachers concerning the evaluation process and its contribution to teacher professional growth?

The answer to this question was gained through obtaining frequencies. As Table 7 indicates, most teachers felt that not a great deal of time was used during the teaching day for professional development. These results are in line with the research of scholars in the field that most teacher evaluation systems did not assist teachers to grow professionally or to highly integrate professional development as a component of teacher evaluation (McGreal, 1983). The research in the review of the literature also indicated that the reason teachers view the process of evaluation as non-productive is due mainly because it is not directly linked to professional development (Shinkfield & Stufflebeam, 1995).

Table 7: Frequency Table on Views of Teachers on the Evaluation Process on Professional Growth

	Frequency	Percent	Cumulative Percent
None	23	29.1	29.1
Somewhat	53	67.1	96.2
Great deal	3	3.8	100

*Research Question Five*

What are the views of teachers on the evaluation process and its relevance in improving PSSA scores?

This question was also answered by obtaining frequencies. The data in the Table 8 suggested that 25% of the teachers did feel that the evaluation process was relevant to improving PSSA scores. Only 7.6% of the teachers did not feel that the evaluation process was used extensively, nor was it relevant, to improve PSSA scores. The main concern was that roughly a quarter of the teachers indicated the evaluation process as neither useful nor relevant in improving PSSA scores. In other words, a significant proportion of teachers saw their evaluation as independent of PSSA scores. This research supports Bastarache's (2000) conclusion that teachers felt that their evaluations had little or no impact on improving the performance of test scores. This research showed that teachers see their evaluation as independent of PSSA scores.

Table 8: Frequency Table on Views of Teachers on the Evaluation Process and Relevance in Improving PSSA Scores

	Frequency	Percent	Cumulative percent
Not considered	20	25.3	25.3
Somewhat	53	67.1	92.4
Used extensively	6	7.6	100

*Research Question Six*

What are the major concerns that teachers have with the current system of teacher evaluation?

Table 9: Descriptive Statistics for Teacher Concerns with the Current Evaluation System

	AYP (n=49) Mean	SD	Non-AYP (n=25) Mean	SD
Form of standards	2.14	.86	2.24	.97
Standards for evaluation	2.60	1.32	2.52	1.48
Examination of student achievement	2.71	1.26	3.00	1.32
Examination of school records	2.96	1.15	3.68	1.25
Number of formal observations per year	2.67	.826	2.80	1.08
Frequency of informal observations	2.29	.935	2.20	1.04
Average length of informal evaluation	2.57	1.17	1.48	.79
Number of different people evaluating	1.56	1.13	1.48	.82
Frequency of formal feedback	2.81	1.19	2.76	1.42
Frequency of informal feedback	2.76	1.15	2.40	1.55
Quality of ideas and suggestions	3.06	1.07	2.80	1.47
Amount of time spent on entire evaluation process	3.06	.99	2.92	.91
Time allotted for professional development	2.63	1.20	2.12	1.09
Available training	2.98	1.15	2.64	1.04
Intended role of evaluation	2.94	1.21	3.04	1.21
Impact of collective bargaining agreement	3.00	1.16	2.67	1.34

This question was answered by obtaining the mean for each item. Table 9 displays the results. Items with small means indicated that they were rated low by the teachers, an indication of a concern. These numbers indicated that teachers in AYP districts were generally less concerned about the evaluation process overall and, specifically, were concerned to a greater extent about factors not clearly related to student achievement. In particular, teachers in AYP districts were concerned about the quality of

ideas and suggestions, the amount of time the evaluation consumes, and the impact on the collective bargaining agreement. Teachers in non-AYP districts were relatively more concerned with factors more directly related to student achievement, such as examination of student achievement and examination of school records. Teachers in such districts were also more concerned with the intended role of the evaluation, probably relating to the riskier situation in non-AYP districts. The chart also indicated that both AYP and non-AYP schools had a top concern with the number of different people evaluating them.

## Summary

Overall, the views of teachers toward formal evaluation were positive. The perceptions of teachers on the evaluator were also positive. Four of five teachers revealed that they think that their evaluator is credible, helpful, trustworthy, and non-threatening. Attitudes of teachers did differ depending on the school's AYP status; however, significant differences existed on only a few items. In non-AYP schools, the results showed that teachers believed the standards were not endorsed in their evaluation as much as they were in non-AYP schools. Non-AYP schools' teachers perceived that they had evaluators who spent more time examining class records than the teachers whose schools met AYP. In both AYP and non-AYP schools, the teachers felt insufficient time was devoted to professional development.

## CHAPTER FIVE: DISCUSSION

A great emphasis has been placed on teacher accountability with the emergence of NCLB in 2001. Key provisions of NCLB establish that the quality of the school staff has a direct correlation to student academic performance and that teacher quality does have a tremendous influence on student achievement and is a primary factor in determining the success of schools (U.S. Department of Education, 2004).

Not until recently has there been a move to change the way the teacher evaluation process, the primary means of identifying the quality of teachers, is conducted in school districts. For teacher evaluation to be effective and meet the demands of the 21<sup>st</sup> century, all system participants must develop shared goals. Obstacles to effective evaluation lie quite often in the attitudes of teachers and evaluators about each other, about consistency, feedback, and multiple evaluative purposes (Danielson, 2000).

The purpose of this study was to understand the relationship between teachers' attitudes toward teacher evaluation in high- and low-achieving middle schools based on the PSSA. The findings of this study are based on data collected with the TEP instrument.

### Summary of Findings

A total of 400 surveys were sent, and 79 were returned, resulting in a 20% response rate. A modified version of the TEP was used to gather data. Over 65% of the participants came from schools that met AYP. The highest percentage (38%) of the

teachers participating in the study had teaching experience in the range of 6-10 years. Finally, 61% of the respondents were female.

The first research question examined the views of teachers towards formal evaluation. The hypothesis was that there were negative attitudes of teachers towards teacher evaluation. Mean scores were calculated to measure perceptions of teachers on the evaluator. The data suggested that, overall; teachers had a positive perception of the evaluator. Teachers claimed that their evaluators are credible, trustworthy, and non-threatening. These findings rejected the hypothesis that teachers were negative towards teacher evaluation.

To examine research questions two and three, independent t-tests were run to determine whether teachers' attitudes on teacher evaluation differed depending upon the designation of a school's AYP status within the district in which they work. The hypothesis was that there were significant differences in teacher attitudes towards teacher evaluation between schools that met AYP and those that did not meet AYP. The attitudes differed on most of the items; however, significant differences were noted on only three particular items in the questionnaire. Differences occurred between the AYP and non-AYP schools regarding endorsement of standards, examination of classroom records, and length of informal observations. Teachers from non-AYP schools perceived that their evaluators spent more time examining classroom records as opposed to those from AYP schools. Therefore, the findings supported the hypothesis that there were significant differences in teacher attitudes towards teacher evaluation between schools that met AYP and those schools that did not.

The third research question asked if there were any significant differences in teacher perceptions of teacher evaluation based on sex, age, years of teaching and degree level. The hypothesis was that there were significant differences in teacher perceptions of teacher evaluation based on sex, age, and teaching experience. A significant positive relationship exists between age and perception. As age increased, the more senior teachers tended to rate their teaching competence higher. Teachers with zero to five years of teaching experience rated their competency lower than teachers with more years of experience. A post-hoc Tukey test revealed that most of the differences were between the least experienced group with zero to five years of experience compared to all other groups; nevertheless, but no significant differences were found among other groups. The findings supported the hypothesis that there were significant differences in teacher perceptions of teacher evaluation based on sex, age, and teacher experience; however, these findings were focused mainly on how teachers viewed themselves in the teaching capacity.

The fourth research question asked what were the views of teachers on the evaluation process in contributing to teacher professional growth. The hypothesis was that teachers viewed the evaluation process as beneficial to teacher professional growth. A majority of the teachers in this study felt that not a great deal of time was used during the teaching day for professional development; therefore, this is contrary to the original hypothesis.

Research question five asked what were the views of teachers on the evaluation process and relevance in improving PSSA scores. The hypothesis was that teachers viewed the evaluation process as relevant to improving PSSA scores. The results of this

study indicated that 25% of the teachers did not feel that the evaluation process was relevant to improving PSSA scores. 7.6% did not feel that the evaluation process was used extensively to improve scores or that it was relevant in teacher observations. The main concern was that roughly a quarter of the teachers indicated that the evaluation process was neither useful nor relevant in improving scores. A significant portion of teachers saw their evaluation as independent of PSSA scores. Therefore, the findings of this study were contrary to the original hypothesis.

The sixth and final research question asked what were the major concerns that teachers had with the current system of teacher evaluation. The original hypothesis was that teachers had no major concerns with the current system of teacher evaluation. The study found that teachers do have general concerns with the current system of teacher evaluation. This study's findings are contrary to the original hypothesis. Teachers in AYP districts were concerned with the quality of ideas and suggestions, the amount of time the evaluation consumed and the impact of collective bargaining units. Teachers in Non-AYP districts were concerned about factors related to student achievement and school records. Teachers in these districts were also more concerned with the intended role of the evaluation. Both schools have a top concern with the number of different people evaluating them.

### Conclusions

In an age when increased accountability and demands for improved achievement challenge public schools, the evaluation of teachers is highly important. Overall findings of this study offer many implications for professional educational practice. Based on specific findings the following conclusions are warranted:

1. Teachers in AYP districts are generally less concerned about the evaluation process overall and, specifically, are concerned to a greater extent about factors not clearly related to student achievement, such as the number of people evaluating them. On the other hand, teachers in non-AYP districts are relatively more concerned with factors more directly related to student achievement, such as school records and the intended role of evaluation.
2. Approximately, one quarter of all teachers view the evaluation process as neither useful nor relevant in improving PSSA scores. In other words, a significant number of teachers see their evaluation as independent of PSSA scores.
3. Views of teachers toward the contribution of the evaluation process to teacher professional growth are not positive. As the results indicate, most teachers felt that insufficient time is used during the teaching day for professional development. As a researcher and practitioner in the field, it is quite often noted that teacher evaluation is not associated with professional development or growth. This is due to the fact that time is not given to district administrators or teachers to do cooperative planning of adult learning activities. Since this research indicated a concern in the area of professional development, it would be wise for administrators and districts to review their professional development plans and tie in activities directly related to teacher evaluations.

Danielson (2000) indicated that obstacles to effective evaluation lie quite often in the attitudes of teachers and evaluators about each other, about consistency, feedback, and multiple purposes used for evaluation. Previous research has shown that teachers

viewed summative teacher evaluation practices as a meaningless exercise as currently practiced in the majority of school districts (Danielson & McGreal, 2000).

This study's data supported the conclusions of the bulk of the scholarly literature, which demonstrated that most teacher evaluation systems did not assist teachers to grow professionally (McGreal, 1983; Sergiovanni & Starrat, 1988; Shinkfield & Stufflebean, 1995). This study's findings have implications for meaningful change in development and implementation of teacher evaluation practices integrated with professional development for teachers. Research indicated that teachers view traditional models of evaluation as non-productive, mainly because traditional models have no link to a teacher's development. Therefore, schools should consider ways to create focused and continuous professional development activities integrated with the activities of the teacher evaluation program.

The data from this study also supported Bastarache's (2000) conclusion that teachers felt that their evaluations had little or no impact on improving their classroom performance or on improving student test scores. This view contradicts research conducted by Turpin (2005) who found that teachers perceive their evaluations as beneficial to continuous improvement and improved test scores. In fact, the research described in this report showed that teachers saw their evaluation as independent of PSSA scores.

The results of this study revealed some interesting, if somewhat mixed conclusions. Overall, the views of teachers toward formal evaluation were positive. The perceptions of teachers on the evaluator were also positive in schools that met and did not meet adequate yearly progress. The hypothesis was that the attitudes of teachers would

be negative in schools that did not meet adequate yearly progress. Surprisingly, in this study, the results showed that there were significant differences on only a few items, not overall as one would expect. Although no specific leadership behaviors were found to influence the attitudes of teachers on teacher evaluation, consideration of principal actions likely to elevate staff attitudes toward evaluation should continue in research and in practice. A savvy principal should use the findings in this research to conceive ways to sharpen his/her observations of teachers so that they are more sensitive to the positive effects of their craft.

#### Recommendations for Further Research

This study's conclusions suggest several recommendations for future research. Based on previous research and this study's findings, the following recommendations are proposed for further study:

1. The study can be replicated in the entire state to gather a larger sample size.
2. The study can be replicated and completed by high school and elementary teachers to determine if perceptions differ according to grade level.
3. Research can be conducted focusing on those teachers receiving unsatisfactory ratings to view their perception of the evaluation process.
4. Using the Pennsylvania Value-Added Assessment System (PVAAS), a researcher can survey teachers in schools demonstrating progress in test scores to examine if specific teacher perception leads to this improvement.
5. A study can be conducted comparing teacher perceptions and principal perceptions to analyze consistencies and discrepancies between teacher and administrator perceptions.

6. It would be interesting to determine the effects of contract negotiations and teacher union activities upon teachers' perceptions of teacher evaluation in unionized, highly unionized and non-unionized schools.
7. A study can be conducted to determine if teacher evaluation is perceived differently by other stakeholders in the school districts, such as parents, students or board members. Since many parents and other community members are so critical of the performance of public education and NCLB, it would be of great benefit to the educational community to know what these stakeholders think about current teacher evaluation processes.

#### Summary

Scholarly literature review disclosed the need to develop collaborative practices in teacher evaluation focusing on professionalism and collegiality. Future research should focus on multiple sources of data so as to provide the evaluator with a complete picture of the evaluatee's performance. This research study was worthwhile because it uncovered various dynamics related to the process of teacher evaluation and the perceptions teachers have regarding it. As a practitioner in the field, teacher evaluation has been one of the most confusing, difficult job responsibilities that I have faced as an administrator.

During the process of completing this body of work, I have learned that although there are numerous studies and research on teacher evaluation, there is not much research on the attitudes of teachers towards it. Most of the research was dedicated to teacher evaluation and student achievement, but studies I reviewed on teacher attitudes towards evaluation all mentioned a need for professional development. I am not terribly surprised by that finding. In my experience as an administrator, teachers typically look for ways to

enhance their skills and professional development. I have also witnessed first-hand the distrust that teachers have regarding their evaluations. My research for this study further confirmed that teachers view professional evaluation as a punitive process. This is due to the fact that principals typically only visit classrooms a few times per year and that teachers tend to resent a process that they perceive to be a waste of time, or not a true assessment of what goes on in the classroom on a daily basis. As this study points out, if researchers can pinpoint strengths and weaknesses in the process of teacher evaluation, then administrators can use that information more effectively to enhance a teacher's skills and, ultimately, to help in increasing student achievement. In the end, that's the most ambitious goal we can have as educators.

## References

- Andrews, R. & Soder, R. (1987). Principal leadership and student achievement. *Educational Leadership*, 44(6), 9.
- Argotsinger, J. (2002). *Components of performance-based teacher evaluation systems related to teacher growth: A case study*. Doctoral dissertation, University of Missouri, Columbia.
- Atkins, A. O. (1996). *Teachers' opinions of the evaluation process*. (ERIC Document Reproduction Service No. ED398628)
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.
- Barling, J. & Beattie, R. (1983). Self-efficacy beliefs and sales performance. *Journal of Organizational Behavior Management*, 5, 41-51.
- Bastarache, R. (2000). *Purposes, methods, and effectiveness of teacher evaluation: Perceptions of urban elementary teachers and principals*. Doctoral dissertation, University of Massachusetts.
- Battaglia, M. (2001). *Teacher evaluation systems in southwestern Pennsylvania emanating from restructuring process*. Doctoral dissertation, University of Pittsburgh, PA.
- Beecher, D. E. (1949). *The evaluation of teaching: Backgrounds and concepts*. Syracuse, NY: Syracuse University Press.

- Bolin, F. S. & Panaritis, P. (1992). Searching for a common purpose: A perspective on the history of supervision. In C.D. Glickman (Ed.), *Supervision in Transition* (pp. 30-43). Alexandria, VA: Association for Supervision and Curriculum Development.
- Bosswell, W. & Boudreau, J. (2000). Employee satisfaction with performance appraisals and appraisers: The role of perceived appraisal use. *Human Resource Development Quarterly*, 11(3).
- Burchfield, M. (1986). *A comparison of principals' and teachers' perceptions of teacher evaluation*. Doctoral dissertation, University of Alabama, Birmingham.
- Castetter, W. B. (1981). *The personnel function in educational administration* (3rd ed.). New York: MacMillan.
- Catano, N. & Stronge, J. H. (2006). What are principals expected to do? Congruence between principal, evaluation and performance standards. *NASSP Bulletin*, 90(3), 221-237.
- Charters, W. W. & Waples, D. (1929). *The commonwealth teacher training study*. Chicago: University of Chicago Press.
- Christen, W. & Murphy, T. (1987). In-service training and peer evaluation: An integrated program for faculty development. *NASSP Bulletin*, 71(500), 10, 12-18.
- Clark, D. (1993). *Teacher evaluation: A review of the literature with implications for educators*. (ERIC Document Reproduction Service No. ED359174)
- Colby, S. (2002, April). *Teacher evaluation: A review of the literature*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. & McGreal, T. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Darling-Hammond, L. (1986). A proposal for evaluation in the teaching profession. *Elementary School Journal*, 86(4), 531-551.
- Darling-Hammond, L. (1990). Teacher evaluation in transition: Emerging roles and evolving methods. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 17-34). Newbury Park, CA: SAGE Publications.
- Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*. New York: National Commission on Teaching and America's Future.
- Darling-Hammond, L. & Ball, D. (1997). *Teaching for high standards: What policymakers need to know and be able to do*. Paper presented for the National Education Goals Panel, Washington, D.C.
- Darling-Hammond, L.; Wise, A.; & Klein, S. (1999). *A license to teach: Raising standards for teaching*. San Francisco: Jossey-Bass.
- Darling-Hammond, L.; Wise, A.; & Pease, S. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.

- Duke, D. (1987). *School leadership and instructional improvement*. New York: Random House.
- Duke, D. (1990). Developing teacher evaluation systems that promote professional growth. *Journal of Personnel Evaluation in Education*, 4(2), 131-144.
- Duke, D. & Stiggins, R. J. (1986). *Teacher evaluation: Five keys to growth*. Washington, DC: National Institution of Education. (ERIC Document Reproduction Service No. ED275069)
- Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-18.
- Education Commission of the States. (1984). Task force on education for economic growth. *Journal of Children in Contemporary Society*, 16(3-4), 11-47.
- Ellett, C. D. & Garland, J. S. (1987). Teacher evaluation practices in our largest school districts: Are they living up to “state of the art” systems? *Journal of Personnel Evaluation in Education*, 1, 69-92.
- Elsbree, W. (1939). *The American teacher: Evolution of a profession in a democracy*. New York: American Book Company.
- Fowler, S. (2001). *Franklin regional growth-oriented teacher evaluation plan opportunities and obstacles: An analysis of teachers’ perceptions of their first year’s experiences in a growth-oriented teacher evaluation plan*. Doctoral dissertation, Duquesne University, Pittsburgh, PA.
- Frase, L. & Streshly, W. (1994). Lack of accuracy, feedback and commitment in teacher evaluation. *Journal of Personnel Evaluation in Education*, 8(1), 47-57.

- Frels, K.; Cooper, T. T.; & Reagan, B. R. (1986). *Practical aspects of teacher evaluation*. New York: National Organization on Legal Problems of Teaching.
- Gallagher, H. A. (2002). *The relationship between measures of teacher quality and student achievement: The case of Vaughn Elementary School*. Unpublished doctoral dissertation, University of Wisconsin, Madison.
- Gibb, C. (1989). *Perceptions of classroom teachers and supervisory principals relating to teacher evaluation policies and processes*. Unpublished doctoral dissertation, University of Southern California, Berkley.
- Gist, M. (1987). Self efficacy: Implications for organizational behavior and human resource management. *The Academy of Management Review*, 12(3), 472-485.
- Glasser, W. (1985). *Control theory: A new explanation of how we control our lives*. New York: Harper & Row.
- Goldhaber, D. D.; Brewer, D. J.; & Anderson, D. (1999). A three-way error components analysis of educational productivity. *Education Economics*, 7(3).
- Haefele, D. (1992). Evaluating teachers: An alternative method. *Journal of Personnel Evaluation in Education*, 5(4), 335-345.
- Haefele, D. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education*, 7(1), 21-31.
- Haney, W.; Madaus, G.; & Kreitzer, A. (1987). Charms talismanic: Testing teachers for the improvement of American education. In E. Rothkopf (Ed.), *Review of research in education* (pp. 169-238). Washington, D.C.: American Educational Research Association.
- Harris, B. M. (1986). *Developmental teacher evaluation*. Boston: Allyn and Bacon.

- Hutto, R. (2001). *Teacher evaluation and development and student performance*. Unpublished doctoral dissertation, University of Texas, Austin.
- Karier, C. (1982). Supervision in historic perspective. In T. J. Sergiovanni (Ed.), *Supervision of teaching* (pp. 2-15). Washington, D.C.: Association for Supervision and Curriculum Development.
- Kent, C. C. (2002). *A school practitioner guide: An English translation of the No Child Left Behind Act of 2001: The challenges and opportunities to produce literate students by 2013-2014* (5 ed.). University of Pittsburgh: Tri-State Area School Study Council.
- Lawler, E. (1969). Job design and employee motivation. *Personnel Psychology*, 22, 426-435.
- Leithwood, K. & Steinbach, R. (1994). *Expert problem solving: Evidence from school and school districts leaders*. Albany, NY: State University of New York Press.
- Loup, K.; Garland, J.; Ellett, C.; & Rugatt, J. (1996). Ten years later: Findings from the replication of a study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education*, 10(3), 203-226.
- Madaus, G. F. & Pullin, D. (1987). Teacher certification tests: Do they really measure what we need to know? *Phi Delta Kappan*, 69, 31-38.
- McGreal, T. L. (1982). Effective teacher evaluation systems. *Educational Leadership*, 39(4), 303-305.
- McGreal, T. L. (1983). *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.

- McGreal, T. L. (1988). Evaluation for enhancing instruction: Liking teacher evaluation and staff development. In S. J. Stanley and J. Popham (Eds.), *Teacher evaluation and staff development: Six prescriptions for success* (pp. 13-16). Alexandria, VA: Association for Supervision and Curriculum Development.
- McGreal, T. L. (Speaker). (1994). *The next generation of teacher evaluation*. (Cassette Recording No. 94-4625). Alexandria, VA: Association for Supervision and Curriculum Development.
- McLaughlin, M. W. & Shepard, L. A. (1995). *Improving education through standards-based reform—A report by the National Academy of Education Panel on standards-based education reform*. Washington, DC: National Academy of Education.
- Medley, D. M.; Coker, H.; & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach*. New York: Longman.
- Milanowski, A. T.; Kimball, S. M.; & White, B. (2004, April). *The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites*. Paper presented at the Consortium for Policy Research in Education, San Diego, CA.
- Millman, J. (1981). Student achievement as a measure of teacher competence. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 146-166). Beverly Hills, CA: SAGE Publications.
- Millman, J. & Darling-Hammond, L. (Eds.). (1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: SAGE Publications.

- Mitchell, Jr., J. V.; Jr., Wise, S. L.; & Plake, B. S. (Eds.). (1990). *Assessment of teaching: Purposes, practices, and implications for the profession*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- National Center for Education Statistics. (1994). *Data compendium for the NAEP 1992 reading assessment of the nation and the states*. Washington, D.C.: U.S. Department of Education.
- National Commission on Excellence in Education Report. (1983). *A nation at risk: The imperative for educational reform*. Washington, D.C.: United States Government Printing.
- National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York: National Commission on Teaching and America's Future.
- Natriello, G. (1984). Teacher's perceptions of the frequency of evaluation and assessments of their effort and effectiveness. *American Educational Research Journal*, 21(3), 579-595.
- Natriello, G. (1990). Intended and unintended consequences: Purposes and effects of teacher evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation* (pp. 35-45). Newbury Park, CA: SAGE Publications.
- Nelson, B. & Sassi, A. (2000). Shifting approaches to supervision: The case of mathematics supervision. *Educational Administration Quarterly*, 36(4), 553-584.
- Newton, E. H. & Braithwaite, W. E. (1988). Teacher perspectives on the evaluation of teachers. *Educational Studies*, 14(3), 275-88.

- Odden, A. & Kelley, C. (2002). *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools* (2nd ed.). Thousand Oaks, CA: Corwin.
- Pennsylvania Accountability System*. Retrieved November 20, 2006, from <http://www.pde.state.pa.us/pas/site/default.asp>
- Peterson, C. H. (1982). *A century's growth in teacher evaluation in the United States*. New York: Vantage Press.
- Peterson, K. (1987). Teacher evaluation with multiple and variable lines of evidence. *American Educational Research Journal*, 24(2), 311-317.
- Peterson, K. (1995). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin.
- Popham, W. (1986). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education*, 7(1).
- Porter, S. R. & Whitcomb, M. E. (2003). The Impact of Lottery Incentives on Student Response Rates. *Research in Higher Education*, 44(4).
- Robinson, G. E. (1983). *Paying teachers for performance and productivity: Learning from experience*. Arlington, VA: ERS Concerns in Education. (ERIC Document Reproduction Service No. ED233483)
- Schmoker, M. (2006). *Results now: How we can achieve unprecedented improvements in teaching and learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Scriven, M. (1981). Summative Teacher Evaluation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary*

- and secondary school teachers* (pp. 244-271). Newbury Park, CA: SAGE Publications.
- Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education*, 8(2), 151-184.
- Sergiovanni, T. J. (1991). *The principalship: A reflective practice perspective* (2nd ed.). Boston: Allyn and Bacon.
- Sergiovanni, T. J. & Starrat, R. J. (1988). *Supervision: Human perspectives*. New York: McGraw-Hill.
- Shinkfield, A. J. & Stufflebeam, D. L. (1995). *Teacher evaluation: Guide to effective practice*. Boston: Kluwer Academic Publishers.
- Spiggle, J. (2003). *Relationship between teacher performance and student growth outcomes in a school district in a North Carolina public school 5<sup>th</sup> grade*. Unpublished doctoral dissertation, University of North Carolina, Charlotte.
- Stiggins, R. J. & Bridgeford, N. J. (1984). *Performance assessment for teacher development*. Washington, DC: National Institution of Education. (ERIC Document Reproduction Service No. ED242717)
- Stiggins, R. J. & Duke, D. (1988). *The case for commitment to teacher growth: Research on teacher evaluation*. Albany, NY: State University of New York Press.
- Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher*, 13(9), 11-18.
- Strike, K. & Bull, B. (1981). *Fairness and the legal context of teacher evaluation handbook of teacher evaluation*. Beverly Hill, CA: SAGE Publications.

- Stufflebeam, D. L. (1992, April 24). *Competing rationales and associated models and approaches for evaluating the performance of teachers*. Presented at the annual meeting of the American Educational Research Association, San Francisco.
- Success for All Foundation*. Retrieved November 8, 2006, from <http://successforall.com/>
- Tomaino, S. (1997). *The effect of stakeholder position in the school district on perceptions held toward the purpose of teacher evaluation*. Doctoral dissertation, Duquesne University, Pittsburgh, PA.
- Travers, R. M. W. (1981). Criteria of good teaching. In J. Millman (Ed.), *Handbook of Teacher Evaluation*. Beverly Hills, CA: SAGE Publications.
- Turpin, S. (2005). *Teacher's perceptions of the teacher appraisal process*. Doctoral dissertation, Clemson University, South Carolina.
- U.S. Congress. (1993). Goals 2000: Educate America Act. Report from the committee on education and labor, together with dissenting and supplemental dissenting views [to Accompany H.R. 1804]. House of Representatives, 103d Congress, 1st Session. Goals 2000: Educate America Act. (ERIC Document Reproduction Services No. ED361834)
- U.S. Department of Education. (2004). *No Child Left Behind: A toolkit for teachers* (2nd ed.) [Booklet]. Washington, D.C.: Education Publications.
- Weber, J. R. (1987). *Teacher evaluation as a strategy for improving instruction*. *Synthesis of Literature*. Elmhurst, IL: North Central Regional Educational Lab. (ERIC Document Reproduction Service No. ED287213)

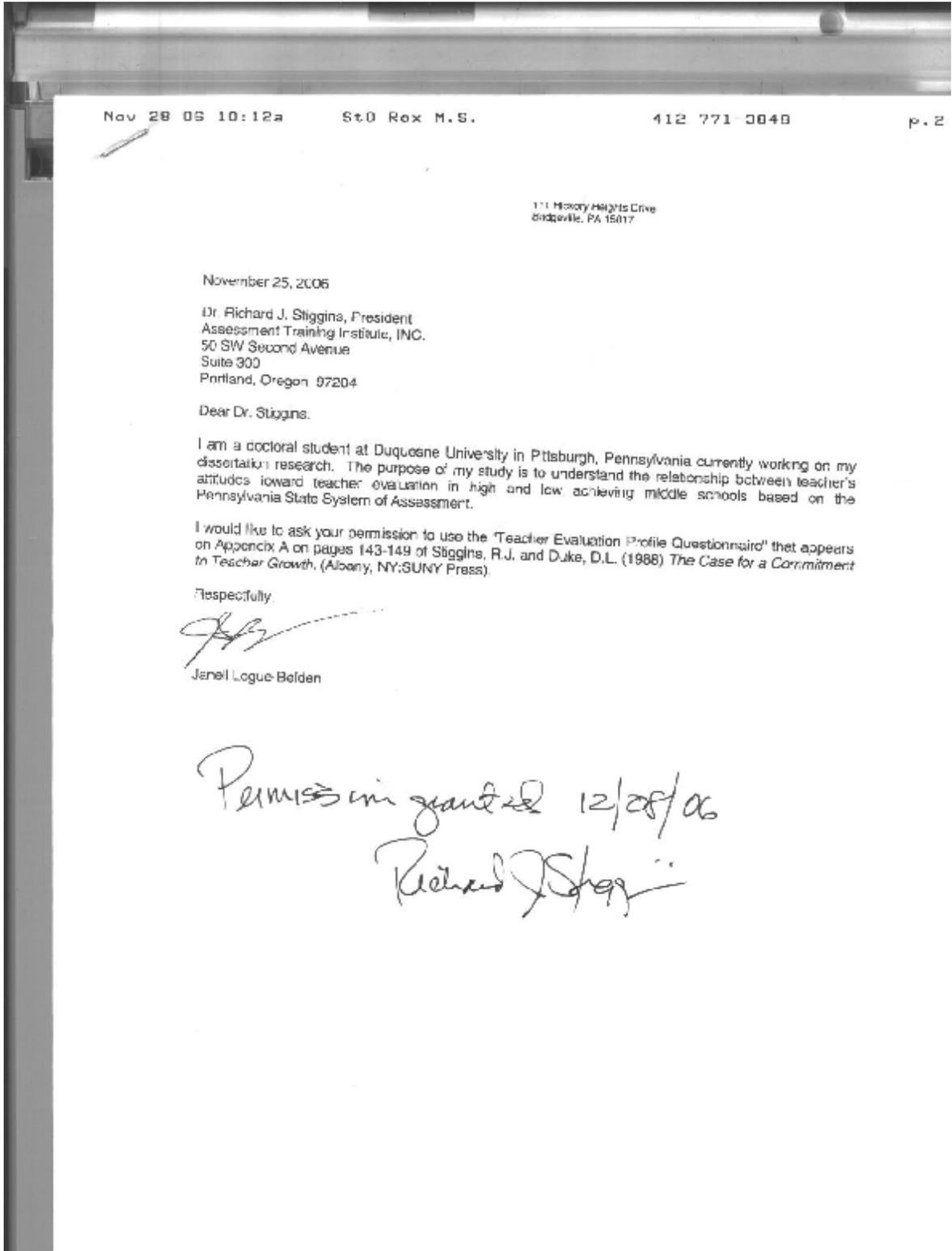
Wise, A. E.; Darling-Hammond, L.; McLaughlin, M. W.; & Bernstein, H. T. (1984).  
*Teacher evaluation: A study of effective practices*. (Report No. R-3139-NIE).  
Santa Monica, CA: The Rand Corporation.

Wood, C. J. & Pohland, P. A. (1979, April). *Teacher evaluation: Organizational maintenance versus stimulation of improved teacher performance. The New Mexico principalship study*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Woolfolk, A. & Hoy, W. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology*, 82(1), 81-91.

## Appendices

Appendix A



Nov 28 06 10:12a

St.O. Rex M.S.

412 771 0848

p. 2

111 Hickory Heights Drive  
Erdreaville, PA 15017

November 25, 2006

Dr. Richard J. Stiggins, President  
Assessment Training Institute, INC.  
50 SW Second Avenue  
Suite 300  
Portland, Oregon 97204

Dear Dr. Stiggins,

I am a doctoral student at Duquesne University in Pittsburgh, Pennsylvania currently working on my dissertation research. The purpose of my study is to understand the relationship between teacher's attitudes toward teacher evaluation in high and low achieving middle schools based on the Pennsylvania State System of Assessment.

I would like to ask your permission to use the 'Teacher Evaluation Profile Questionnaire' that appears on Appendix A on pages 143-149 of Stiggins, R.J. and Duke, D.L. (1986) *The Case for a Commitment to Teacher Growth*. (Albany, NY: SUNY Press).

Respectfully,

Jenell Logue-Belden

Permission granted 12/28/06  
Richard J. Stiggins

*Appendix B*

**Teacher Evaluation Profile (TEP) Questionnaire**

This form has been designed to allow you to describe your experience with teacher evaluation in some detail. Your responses will be combined with those of other teachers to yield a clearer picture of the key ingredients in an effective teacher evaluation experience. The goal of this research is to determine if and how the evaluation process can be revised to help it serve relevant and useful purposes. If we are to reach this goal, it will be important for you to provide frank and honest responses. This is why your answers will remain anonymous.

The process leading toward the end of the year evaluation is expected to consist of formal/informal classroom visits and observations, and conferences or discussions between the teacher and supervisor at least four times over the course of the school year for non-tenured teachers and twice per year for tenured teachers. When reference is made in this questionnaire to teacher evaluation, it should be understood to encompass all of these elements.

**Specific Instructions**

With this definition of teacher evaluation (classroom visits/observations, teacher/supervisor conferences), please reflect on your most recent experience with a teacher evaluation system.

What is your gender?

1. Male
2. Female

What is your age?

\_\_\_\_\_

How many years of teaching experience?

1. 0-5 year
2. 6-10 year
3. 11-15 years
4. 16-20 years
5. 21 +

What is your degree level?

1. Bachelors Degree
2. Masters Degree
3. Doctorate Degree

Did your middle school meet AYP in 2006

1. Yes
2. No
3. Don't know

**As you think about the evaluation experience, how would you rate the overall quality of the evaluation? (1=Waste of time, 5=Helpful)**

*Describe These Attributes of You as a Teacher*

Rate your overall competence as a teacher

*I am minimally competent* 1      2      3      4      5      *I am outstanding*

Rate the strength of your professional expectations of yourself

*I demand little*              1      2      3      4      5      *I demand much*

*Describe your interpersonal manner:*

Orientation to risk taking

*I avoid risks*              1      2      3      4      5      *I take risks*

Orientation of others

*I'm reserved and private* 1      2      3      4      5      *I'm open, public*

Attribution of reasons for your success/failure

*I hold others responsible* 1      2      3      4      5      *I hold myself responsible*

Orientation to change

*I'm slow to change*              1      2      3      4      5      *I'm flexible*

Orientation to experimentation in the classroom

*I don't experiment*              1      2      3      4      5      *I experiment often*

Openness to criticism

*I'm relatively closed*              1      2      3      4      5      *I'm relatively open*

Knowledge of technical aspects of teaching

*I know little*              1      2      3      4      5      *I know a great deal*

Knowledge of subject matter

*I know a little*              1      2      3      4      5      *I know a great deal*

*Describe Your Perceptions of the Person Who Evaluated Your Performance (Most Recently)*

Credibility as a source of feedback

*Not credible*              1      2      3      4      5      *Very credible*

Working relationship with you

*Adversary*              1      2      3      4      5      *Helper*

Level of trust

*Not trustworthy*              1      2      3      4      5      *Trustworthy*

Interpersonal manner

*Threatening*              1      2      3      4      5      *Not threatening*

Temperament

*Impatient*              1      2      3      4      5      *Patient*

Flexibility

*Rigid*              1      2      3      4      5      *Flexible*

Knowledge of technical aspects of teaching

<i>Not knowledgeable</i>	1	2	3	4	5	<i>Knowledgeable</i>
Capacity to demonstrate or model needed improvements						
<i>Low</i>	1	2	3	4	5	<i>High</i>
Familiarity with your particular classroom						
<i>Unfamiliar</i>	1	2	3	4	5	<i>Very familiar</i>
Experience in classrooms in general						
<i>Little</i>	1	2	3	4	5	<i>A great deal</i>
Usefulness of suggestions for improvements						
<i>Useless</i>	1	2	3	4	5	<i>Useful</i>
Persuasiveness of rationale for suggestions						
<i>Not persuasive</i>	1	2	3	4	5	<i>Very persuasive</i>

**Describe These Attributes of the Information Gathered on Your Performance During Your Most Recent Evaluation**

*How were the dimensions of your teaching (standards) to be evaluated addressed?*

Were standards communicated to you?

*Not at all*            1        2        3        4        5    *In great detail*

Were standards clear to you?

*Vague*                1        2        3        4        5    *Clear*

Were standards endorsed by you as appropriate for your classroom?

*Not endorsed*        1        2        3        4        5    *Endorsed*

What was the form of the standards?

1. *Goals to be attained*
2. *Personal and/or professional traits to possess*
3. *Both*

Were the standards:

*The same for all teachers?* 1        2        3        4        5    *Unique to you?*

*To what extent were the following sources of performance information tapped as a part of the evaluation?*

Observation of your classroom performance

*Not considered*        1        2        3        4        5    *Used extensively*

Examination of student achievement

*Not considered*        1        2        3        4        5    *Used extensively*

Examination of classroom or school records (lesson plans. Etc.)

*Not considered*        1        2        3        4        5    *Used extensively*

*To what extent were there observations in your classroom?*

(In these terms, formal refers to observations that were pre-announced and were preceded and followed by a conference with the evaluator: informal refers to unannounced drop-in-visits)

Number of formal observations per year (most recent experience)

1. 0
2. 1
3. 2
4. 3
5. 4 or more

Approximate frequency of informal observations (most recent experience)

1. None
2. Less than 1 per month
3. Once per month
4. Once per week
5. Daily

What was the average length of observation (most recent experience)?

Formal

*Brief (few minutes)*      1      2      3      4      5      *Extended (40+ )*

Informal

*rief (few minutes)*      1      2      3      4      5      *Extended (40+ )*

Number of different people observing and evaluating you during the year

1. Supervisor only
2. Supervisor and 1 other person
3. Supervisor and 2 other persons
4. Supervisor and 3 other persons
5. Other

### **Describe These Attributes of the Feedback You Received**

Amount of information received

*None*      1      2      3      4      5      *Great deal*

Frequency of formal feedback

*Infrequent*      1      2      3      4      5      *Frequent*

Frequency of informal feedback

*Infrequent*      1      2      3      4      5      *Frequent*

Depth of information provided

*Shallow*      1      2      3      4      5      *In-depth*

Quality of ideas and suggestions

*Low*      1      2      3      4      5      *High*

Specificity of information provided

*General*      1      2      3      4      5      *Specific*

Nature of information provided

*Judgmental*      1      2      3      4      5      *Descriptive*

Timing of the feedback

*Delayed*      1      2      3      4      5      *Immediate*

Feedback focused on district teaching standards

*Ignored them*      1      2      3      4      5      *Reflected them*



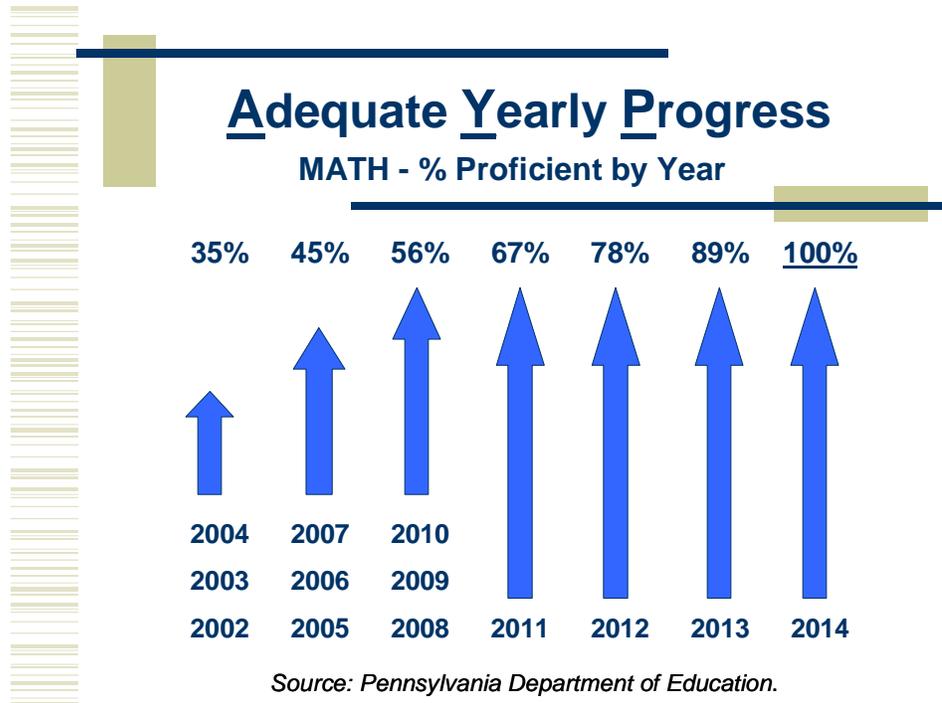


Figure 2: Adequate Yearly Progress (AYP) in Math

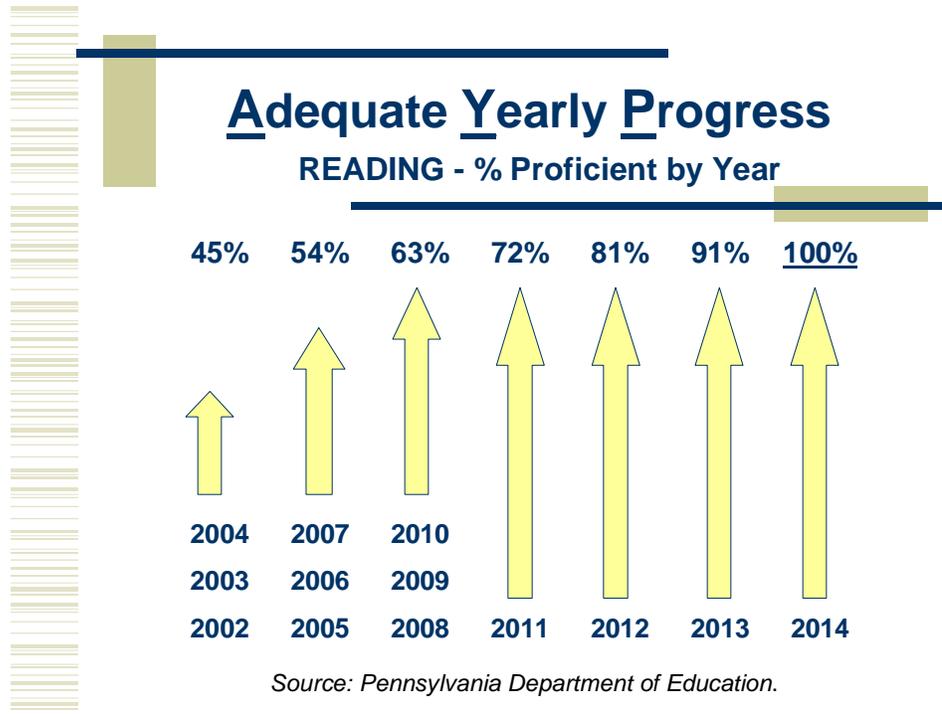


Figure 3: Adequate Yearly Progress (AYP) in Reading