

ABSTRACT

Title of Dissertation: TEACHER PERCEPTIONS OF TEACHER
EVALUATION USING THE TEACHER
PERFORMANCE ASSESSMENT SYSTEM AND
FACTORS THAT CONTRIBUTE TO TEACHER
QUALITY, PROFESSIONAL GROWTH, AND
INSTRUCTIONAL IMPROVEMENT OVER
TIME

Alexander C. Jaffurs, Doctor of Education, 2017

Dissertation Directed by: Patricia M. Richardson, Ph.D
Department of Education

The primary purpose of teacher evaluation is to improve teaching practice, which results in increased student achievement. In practice, however, evaluation systems have been generally used as sorting mechanisms for identifying the lowest performing teachers for selective termination. The school system in this study, like others, aspires to have all of its teachers consistently performing at a highly effective level. The problem of practice faced by the school system is the inability of a large number of teachers rated “effective” to summarily improve their practice over time and move to the “highly effective” rating. In essence, how does a teacher evaluation metric maximize the chances that those who remain in the profession become accomplished practitioners? This research triangulates teacher evaluation, self-reflection and their roles in improving teacher quality. The prevailing thought is that teachers who willingly engage in more formalized self-reflection and self-assessment yield higher degrees of teacher effectiveness

as measured on a local teacher evaluation. The central focus of this study will investigate tenured teachers' perceptions of the effect of their teacher evaluation tool on teacher quality and other factors that contribute to a teacher's improvement of instructional performance over time. The researcher would also like to investigate the extent to which teacher cohorts – differentiated by demographic data - engage in formalized practices of self-reflection about their own teaching practice. Lastly, the researcher would like to determine whether or not tenured teachers who are evaluated with the local teacher evaluation tool actually improve their teacher effectiveness over time.

This study was conducted in a public, K-12 school system with 1420 teachers employed - 39 of which are National Board Certified. This schools system is located in a rural/suburban school system and has utilized its current teacher evaluation system since 2000.

The findings of this study indicated that the majority of teachers – disaggregated by demographic teacher cohort - viewed their local teacher evaluation system somewhere along the continuum of neutral to satisfactory as a tool for building a teacher's effectiveness over time. The overwhelming majority of teachers embraced the post-conference as the most impactful part of the entire evaluation process in building teacher quality; the least impactful was the pre-conference. Additionally, teacher respondents – agnostic of demographic – opined that while the local teacher evaluation system was perceived to be a both quality control and a compliance factor for teachers, less than half of all respondents believe that the system, assists teachers

formatively as a tool for professional development. Per the respondents, it should be noted that the teacher evaluation system elicited the strongest reactions – both positive and negative - in teachers having experienced more than 20 formal observations. The research also conveyed that most teachers reported that there was much more embedded self-reflection in the evaluation system than hypothesized; most prominently, teachers cited that *audio-taping, reviewing student performance data, completing a self-reflective checklist, and engaging in unstructured self-reflection* were a few of the assorted self-reflective activities were facilitated by the evaluation system. Moreover, the data clearly demonstrated that all teachers engage in high degrees of reflection regardless of demographic cohort and a majority of teachers claim to already know how to “self-reflect.” In other words, the highest self-reported degree of reflection were those teachers already rated as “highly effective” in the local evaluation system. A prevalent trend in the data was that degrees of self-reflection matter and build more pronounced levels of teacher effectiveness over time. In essence, the fact that teachers participate in reflection does not seem to impact teacher quality; rather, the degree and amount to which one reflects is actually what matters in building instructional capacity in teachers. Other noticeable trends in the data were as follows: more years of teaching experience was inversely related to the degree to which a teacher self-reflects; over 30% of teachers with more than 20 years of experience reported that they do not self-reflect at all; the non-NBCT teacher cohort out reflects the NBCT cohort; NBCT teachers had the highest average evaluation rating out of every teacher cohort; and, teaching experience seems to mute any lack of reflection in a teacher’s evaluation rating; The

other noticeable trend was that more formal observations for teachers did not translate into higher evaluation ratings over time. Overall, the two most impactful professional development activities cited by teachers were the following: participation in professional learning communities and peer coaching and mentoring, respectively.

**TEACHER PERCEPTIONS OF TEACHER EVALUATION USING
THE TEACHER PERFORMANCE ASSESSMENT SYSTEM AND
FACTORS THAT CONTRIBUTE TO TEACHER QUALITY,
PROFESSIONAL GROWTH, AND INSTRUCTIONAL
IMPROVEMENT OVER TIME**

By:

Alexander C. Jaffurs

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Education

2017

Advisory Committee:

Patricia M. Richardson, Chair

Helene Cohen

Margaret J. McLaughlin, Co-chair

Olivia Saracho

Wayne Slater, Dean's Representative

© Copyright by
Alexander C. Jaffurs
2017

Dedication

I would like to dedicate this dissertation to my family, immediate and extended, that all made this possible and, especially, to my wife, Christine; she has been the apple of my eye since 1989. To all of my former students and teachers, I appreciate all of the insight that you have given me over the years...from which I learned a great deal and more than you know.

Acknowledgements

First and foremost, I would like to acknowledge my advisor, Dr. Patricia Richardson, of whom I have known for close to two decades. Thanks as well to Dr. Margaret McLaughlin and all of our UMD professors who have been consummate professionals throughout this process. Also, I would be remiss if I did not acknowledge my high school calculus teacher, the late Ross Scarcelli, who taught me all about “teeth and eye balls” and saved me from the inglorious fate of practicing law. Additionally, my first mentor in 1994 - Sandy Laughlin – who taught me that relationships with students were most important for teaching; the content a distant second. Last to acknowledge - yet always first in my mind - are my late parents who sacrificed everything for their children and were the greatest and most selfless people I have ever met.

TABLE OF CONTENTS

Dedication.....	iii
Acknowledgments.....	iv
List of Tables.....	vii
List of Figures.....	viii
List of Acronyms.....	x
SECTION I: INTRODUCTION.....	1
Introduction to the Problem	1
Purpose of the Study	5
The Problem	6
Literature Review	8
The Evolution in Teacher Evaluation	8
The Shift in Teacher Evaluation.....	9
Teacher Evaluation in Maryland	13
Current Criticisms of the Teacher Evaluation Process	16
Improving the Teacher Evaluation Process	20
Effective Teacher Evaluation and Professional Growth.....	23
Self-Assessment and the National Board Certification Process	35
Fostering Reflection and Self-Assessment	36
The Five Most Prominent Models of Teacher Evaluation.....	40
Prior Attempts to Solve the Problem	43
Characteristics of the TPAS	45
Limitations of TPAS	46
Summary of Literature Review	51
Proposed Investigation	54
SECTION II: METHODOLOGY.....	55
Research Questions.....	57
Research Design.....	58
Design and Methods.....	61
Pretesting the instruments	61
Instrument.....	62
Limitations of the Study	62
Conceptual Framework.....	64
Research Setting.....	65
Population and Sample.....	65
Survey Pilot Test.....	68
Procedures/Data Collection Methods	68
Hypothesis	73
Human Subject and Review	73
Summary	75

SECTION III: RESULTS.....	77
Introduction.....	77
Results.....	78
Demographic Data	79
Years of Teaching Experience.....	80
Highest Degree Conferred	81
NBCT vs. non-NBCT Cohort	82
Leveled Teaching Assignment	83
Total Number of Career TPAS Observations	84
Initial and Most Recent TPAS Rating.....	85
Research Question One: Teacher Perception of TPAS	88
Q17: TPAS as a Tool for Teacher Evaluation	93
Q18: The Overall Impact of TPAS	94
Q19: The Most Impactful Part of the TPAS Process.....	97
Q23: Perception of TPAS with Respect to Professional Duties.....	100
Q21 and Q22: Past Perceptions and Future Possibilities of TPAS	103
Research Question Two: Teacher Perception of TPAS Component	
Domains.....	106
Q14, Q15, and Q16: Most and Least Impactful TPAS Domains	106
Research Question Three: Are Teachers Self-Reflective Practitioners?	111
Q20: TPAS Process and Self-Reflection	112
Q27: Impact of Self-Reflection.....	115
Q28: Participation in Self-Reflection.....	118
Q29: Do Teachers Know How to Self-Reflect?	121
Research Question Four: Factors that Impact Teacher Effectiveness.....	124
The Total Number of Formal Observations and the Last TPAS Rating	125
Years of Service and the Last TPAS Rating	127
NBCT Certification and the Last TPAS Rating.....	128
Frequency of Unstructured Reflection and the Last TPAS Rating	128
Frequency of Reflection, NBCT and the Last TPAS Rating	131
Frequency of Reflection, Experience Level and the Last TPAS Rating	133
Frequency of Reflection, Leveled Teacher and the Last TPAS Rating	134
Specific Types of Professional Development for Building Teacher	
Effectiveness	136
Conclusion -Resolution of the Research Questions	137
Recommendations for the District.....	148
Suggestions for Future Research.....	151
Appendices.....	154
Glossary	155
References.....	173

List of Tables

Table 1	Teachers' Responses to Survey Question 23 (by percent)	13
Table 2	Compartmentalizing Teacher's Responses to Q23 (in percent)	15
Table 3	Teachers' Responses to Survey Question 18 (in percent)	21
Table 4	Teachers' Responses to Survey Question 19 (in percent)	29
Table 5	Characterizations of Past and Future TPAS Experiences (in percent)	37
Table 6	Teachers' Responses to Survey Question 14 (in percent) ..	38
Table 7	Teachers' Responses to Survey Question 20 (in percent).....	44
Table 8	Teachers' Responses to Survey Question 29 (in percent)	48
Table 9	Number of Observations and Corresponding TPAS Ratings	49
Table 10	Frequency of Teacher Reflection and Average TPAS Ratings	53
Table 11	Differentiated Frequency of Teacher Reflection and TPAS Ratings	44
Table 12	Frequency of NBCT Reflection and Average TPAS Ratings	48
Table 13	Years of Experience, Frequency of Reflection and Average TPAS Ratings..	49
Table 14	Leveled Teacher, Frequency of Reflection and Average TPAS Ratings	53

List of Figures

- Figure 1. Timeline of Relevant Educator Effectiveness Policy (with respect to Teacher Evaluation)
- Figure 2. The Differences between the Danielson Framework and the National Board Model
- Figure 3. Components of the Five Models of Teacher Evaluation
- Figure 4. Timeline of Relevant Educator Effectiveness Policy (with respect to Teacher Evaluation)
- Figure 5. The Differences between the Danielson Framework and the National Board Model
- Figure 6. Crosswalk for Demographic Data and Survey Questions
- Figure 7. Years of Teaching Experience (by respondent percentage)
- Figure 8. Highest Degree Conferred (by respondent percentage)
- Figure 9. NBCT vs. non-NBCT Cohort (by respondent percentage)
- Figure 10. Leveled Teaching Assignment (by respondent percentage)
- Figure 11. Total number of career TPAS observations (by respondent percentage)
- Figure 12. Self-Reported Initial TPAS rating (by respondent percentage)
- Figure 13. Most Recent Self-Reported TPAS rating (by respondent percentage)
- Figure 14. Crosswalk for Research Question One and Survey Questions
- Figure 15. Teacher perception of TPAS as a tool for teacher evaluation
- Figure 16. Results from Survey Question 18
- Figure 17. Results from Survey Question 19
- Figure 18. Results from Survey Question 23
- Figure 19. Additional Results from Survey Question 23

List of Figures (cont.)

Figure 20. Past Teacher Experience with TPAS

Figure 21. TPAS possibilities in the future

Figure 22. Crosswalk for Research Question Two and Survey Questions

Figure 23. Teacher perceptions of least impactful TPAS Domain

Figure 24. Teacher perceptions of most impactful TPAS Domain

Figure 25. Crosswalk for Research Question Three and Survey Questions

Figure 26. Results from Survey Question 20

Figure 27. Results from Survey Question 27

Figure 28. Q27 respondent results of different demographic groups regarding TPAS process

Figure 29. Percentage of teacher participation in self-reflective activities

Figure 30. A comparison of self-reflective activities between NBCT and non-NBCT

Figure 31. A comparison of the frequency of self-reflective activities between NBCT and non-NBCT

Figure 32. Results from Survey Question 29

Figure 33. Crosswalk for Research Question Four and Survey Questions

Figure 34. Total number of TPAS observations and average TPAS rating

Figure 35. Years of service and average TPAS rating

Figure 36. The Differences between the Danielson Framework and the National Board Model

Figure 37. NBCT and average TPAS rating

Figure 37. Frequency of reflection and average TPAS rating

Figure 38. NBCT's degree of reflection and average TPAS rating

Figure 39. Results from Survey Question 29

List of Acronyms

CCSS – Common Core State Standards

COMAR – The Code of Maryland Regulations

ESEA – Elementary and Secondary Act

ESSA – Every Student Succeeds Act

ETS – Educational Testing Service

MSDE – Maryland State Department of Education

NBC – National Board Certified

NBCT – National Board Certified Teacher

NBPTS – National Board of Professional Teaching Standards

NCLB – No Child Left Behind

RTTT – Race to the Top

TPAS – Teacher Performance Assessment System

SECTION I: INTRODUCTION

Introduction to the Problem and Literature Review

“How does a teacher evaluation metric maximize the chances that those who remain in the profession become accomplished practitioners?” (Thorpe, 2014, p.8)

Introduction

Over decades, the role of K-12 teacher evaluation has progressed from being solely focused on *measurement of a teacher* to a model that integrates elements of *teacher improvement* as well (Hinchey, 2010; Jacob & Lefgren, 2008). No longer is an observation rating the sole component to a teacher’s evaluation; rather, part of a comprehensive teacher growth and development model that includes both formative and summative assessment for teachers (Hull, 2013). As states and schools ponder how to do teacher evaluation differently *and* better, right and left leaning educational think tanks have done numerous studies and written articles accordingly as to how to most effectively evaluate teachers.

Metrics such as classroom observations and student achievement data have been increasingly used to assess teachers in new evaluation systems emerging across the country (Kane, Taylor, Tyler, & Wooten, 2011). Researchers in force are now consistently evaluating the strengths and limitations of teacher evaluation systems and the insistence that the components of such systems are reliable, fair, effective, and efficient (Doherty & Jacobs 2013; Darling-Hammond, 2013; Berry, 2010; Peterson & Comeaux, 1990; Finn Jr, Kanstoroom, & Petrilli, 1999).

The single most important in-school factor for student achievement is teacher quality (Marzano, 2003). If we know that good teachers make a difference (Chetty, Friedman, & Rockoff, 2011), it is not surprising that an acute focus of recent educational policy and

procedures is on evaluating teachers. It has been noted that the move toward meaningful teacher evaluation is to assure greater equity in students' access to good teachers (Sablich & Inman, 2014).

It is widely understood that there are vast differences in the quality of teachers as most can attest to the fact that there have been really good, really bad, and decidedly mediocre ones in a student's school experience. Until recently, teachers were deemed qualified, and were compensated, solely according to academic credentials and years of experience; interestingly enough, classroom performance was not necessarily a component thereof (Whitehurst, Chingos, & Lindquist, 2015). In the last decade, researchers have used student achievement data to quantify teacher performance and thereby measure differences in teacher quality (Kane, Taylor, Tyler, & Wooten, 2011; Hull, 2013; Whitehurst, Chingos, & Lindquist, 2015). Among the recent findings is evidence that having a better teacher not only has a substantial impact on students' test scores at the end of the school year, but also increases their chances of attending college and their earnings as adults (Chetty, Friedman, & Rockoff, 2012; Hanushek, 2010).

So why does teacher evaluation matter? Because regardless of how well a program is designed, it is only as effective as the people who implement it (Stronge, 1993). Thus, a conceptually sound, well designed, and properly implemented evaluation system for teachers is an important – indeed, essential - component of an effective school. Despite the fact that proper assessment and evaluation of teachers is fundamental to successful schools and schooling, this key element in school reform is too frequently neglected - due not to the absence of teacher evaluation, but rather to the implementation of poor evaluation systems and poor evaluation practices (Strong, 2006).

According to Marzano and colleagues (Marzano, Frontier, & Livingston, 2014), there are two overarching purposes of teacher evaluation: quality assurance with professional learning. However, the challenge for current systems is merging those two purposes. If, however, the primary purpose of teacher evaluation should be to improve teaching practice to increase student achievement, as some (Darling-Hammond, 2000; Tucker & Strong, 2005; Marzano, 2012), believe, then current, evaluation systems are generally effective only as sorting mechanisms for identifying the lowest performing teachers for selective termination (Taylor & Tyler, 2012). While the teacher evaluation process has the potential to improve instruction and increase student achievement, teacher evaluation models have historically reflected our knowledge of the learning process at that time as well as enacted policy shifts in legislation (Tucker & Strong, 2005; Coggshall, Rasmussen, Colton, Milton, & Jacques, 2012;. Marzano, 2014).

The Obama administration's Race to the Top competitive grant program initiated an unprecedented wave of state teacher-evaluation reform across the country (McGuinn, 2010). To date, most of the scholarly analysis of this activity has focused on the design of the evaluation instruments (Tyler, 2011) and/or the implementation of the new evaluations by districts and schools (Doyle & Han, 2012). Such recent changes in the policy environment have led states and districts to increase the rigor of their teacher evaluation systems by including more frequent observations and/or student test score data. States and districts nationwide began reforming their evaluation systems as early as 2006 and were further spurred to action by federal programs such as the Teacher Incentive Fund and the 2009 Race to the Top grant program (Lacireno-Paquet, Bocala, & Bailey, 2016). As of 2012, 30 states required that evaluations include evidence of student learning, 25 states required

differentiation of teacher ratings into more than two categories (such as professional practice and student growth), and 39 states required annual classroom observations (National Council on Teacher Quality, 2012).

The passage of the federal Every Student Succeeds Act (ESSA) in December 2015 and the accompanying end of federal policies and requirements regarding teacher evaluation provide a natural inflection point for states to review existing systems and recommit to the primary purpose of evaluation: to support teacher growth and development and to act as but one component within robust systems of talent management and instructional improvement (Teacher Evaluation and Support Systems: A Roadmap for Improvement, 2016). States and districts across the nation are at different stages of implementing teacher evaluation systems, and some have already modified their systems or delayed some consequences based on practical or political considerations (Camera, 2014).

Taken together, the challenges of implementing more complex and higher stakes teacher and principal evaluation systems may seem to point toward most states and districts opting out, scrapping their systems and dialing back oversight altogether (Silverman, 2017). Nevertheless, Silverman (2017) purports that countervailing forces such as existing state law and the “gargantuan effort” to repeal and replace new educator evaluation systems that have been in place in most states and districts for at least a few years are likely to hold evaluation systems to a consistent course and point toward modification rather than complete demolition. During the period in which states were vying for Race to the Top (RTTT) grants, numerous state legislatures passed laws requiring more sophisticated evaluation systems. Most of these laws are still in place, and supersede federal flexibility. Moreover, as challenging as it has been to implement, these evaluation systems and their potential for improvement if

maintained and implemented well, is difficult to justify their demise. Competing priorities under ESSA suggest that, while some may elect the nuclear option, most states and local districts will opt to keep working toward evaluation systems that drive better talent management systems, better support for teachers, and better outcomes for kids (Silverman, 2017). Certainly, as teacher evaluation systems are reviewed and redefined, it is important that all evaluation systems reflect current knowledge of best practices and a focus on continuous teacher improvement.

Purpose of the Study

According to Hull (2011), the power of any comprehensive teacher evaluation measure lies not only in simply identifying the few ineffective teachers to fire or the few highly effective teachers to reward with increased compensation, but also in improving the performance of all teachers by providing timely and specific feedback to help them improve their performance. By focusing teacher evaluation on the continuous improvement of all teachers – not just for high stakes personnel decisions – evaluations will have a greater impact on building teachers’ capacity. Danielson (2011) states that evaluation systems must be meaningful for teachers themselves and allow them to glean insight about their own practice.

Interestingly enough, if the goal of teacher evaluation systems is to be more *formative* and meaningful for teachers themselves, school systems must use metrics that are not only rigorous, valid, and reliable, but also engage teachers in those activities that promote learning—namely self-assessment, reflection on practice, and professional conversation with timely, focused feedback (Danielson, 2011). As previously stated, Robert Marzano (2014) describes the challenge as one of merging the two overarching purposes of teacher evaluation: quality assurance with professional learning. According to Marzano, it is not possible to walk

into a classroom and get a full perspective of a teacher's practice from a walkthrough snippet of teaching unless you know exactly what is happening and which segment is occurring.

Marzano confided that he has sat in the back of a classroom and thought, "there's a symphony going on and I'm the only adult seeing it." That level of isolation, he says, has to change for teaching to improve (Marzano, Frontier, & Livingston, 2014).

Marzano insists that administrators must supervise and evaluate the art and science of teaching in a way that keeps attention squarely focused on student learning—rather than on specific instructional strategies— and provides teachers individual flexibility and creativity in utilizing their own unique teaching practices. In fact, Marzano (2014) has qualified his two common beliefs: that frequent observation is good for teacher development and that focusing on high-yield strategies makes it easier to recognize good teaching with the following statements:

- Frequent feedback is good only if it's accurate, and
- High-yield strategies are useful only if the observer is aware of where the teacher is in the scope of the lesson.
- Bottom line: There are no perfect strategies

The practical question is how can we facilitate real improvement in teacher skills through the use of a teacher evaluation system that is affordable and scalable in schools and school systems (Kane, Taylor, Tyler, & Wooten, 2011)?

The Problem

The school system in this study, like others across the nation, aspires to have all of its teachers consistently performing at a highly effective level. A central problem faced by school systems – in general – is the inability to define what exactly "effective" and/or "highly

effective” teaching looks like via a system’s teacher evaluation metric. Tangential to this issue is that most evaluation instruments have a generic, one-size-fits-all tool to quantify teacher performance and fall short of discriminating ways in which evaluators can best support and develop teachers differentiated among different demographic types such as experienced vs. novice teachers (Rogers & Weems, 2010). As a result, there is a large number of teachers rated as “effective” that are unable to summarily improve their practice over time (within a given evaluation framework) and move to the “highly effective” rating. In other words, how does a teacher evaluation metric maximize the chances that those who remain in the profession become accomplished practitioners? (Thorpe, 2014).

A previous study of districts in the early stages of implementing new teacher evaluation systems found a relationship between school professional climate —specifically principal leadership, teacher influence on school policy, trust—and teachers’ support for and fidelity of implementation of the new systems (Riordan, Lacireno- Paquet, Shakman, Bocala, & Chang, 2015). Given the many challenges in implementing an evaluation system, the researcher wanted to explore what factors contribute to both teachers’ perception *and* satisfaction with teacher evaluation tools and assorted processes. Notwithstanding, the study also wanted to consider what factors embedded in the teacher evaluation system, in particular, contribute to higher degrees of teacher quality, professional growth and instructional improvement over time.

Literature Review

This literature review will focus on the role of teacher evaluation in improving teacher practice. More specifically, the researcher will focus on teacher perception and which factors they believe have the greatest opportunity to improve teacher quality and move “effective” teachers to “highly effective.” The review is organized according to the evolution and subsequent shift of teacher evaluation; current criticisms of the teacher evaluation process; ways that school systems can improve teacher evaluation; an analysis of the five most prominent evaluation systems presently utilized in the United States; prior attempts to solve the problem; and a summary of the literature review.

The Evolution of Teacher Evaluation. For decades, teacher evaluations were little more than a bureaucratic exercise that failed to recognize either excellence or mediocrity in teaching (Hull, 2013). The evolution of “sorting mechanism” teacher evaluation metrics can be traced back to the 1700s when clergy were relied upon to provide guidance to and supervision of teachers. As school systems became more complex throughout the mid-1800s, the need for more specialized guidance for teachers gave rise to the principal teacher as leader and a growing awareness of the importance of pedagogy (Marzano, Frontier, & Livingstone, 2014).

From the late 1800s until right before World War II, educational systems were driven by the notion of scientific management and its two competing views of education (Marzano, Frontier, & Livingstone, 2014). It was during this time that the scientific approach gained strength and acceptance and when data was first used to make decisions on future actions. The period after World War II saw a swing away from the scientific approach to an emphasis on developing the teacher as an individual. This period also saw a proliferation of the responsibilities of the supervisor and whose responsibilities were defined in very specific terms (Marzano, Frontier, &

Livingstone, 2014). One major takeaway of this era was the acknowledgement that administrators could help to facilitate more effective teaching via classroom observations.

Throughout the next era, one of the most influential movements in supervision and evaluation - clinical supervision - was prevalent from the late 1960s to the early 1970s (Marzano, Frontier, & Livingstone, 2014). It was during this time that the Madeline Hunter model was combined with clinical supervision to produce a widely used but often a prescriptive approach to supervision. Developmental and reflective models that were much less prescriptive followed this period. In the 1980's, the RAND study provided a realistic look at the actual practice of supervision and evaluation in districts and schools and concluded that teachers preferred specific as opposed to general feedback (Wise et al., 1984). This realization laid the foundation in the mid-1990s for the popularity of the Charlotte Danielson's Framework of Teaching (2000) to take hold. This model was widely applied through K-12 as the evaluation model of choice for all teachers and is still one of the five most prominent teacher evaluation models of choice in school districts across the United States (ASCD, 2014). In addition to the Danielson Framework, the other most prevalent models are as follows:

- Kim Marshall's *Teacher Evaluation Rubric*;
- Robert Marzano's *Teacher Evaluation Model*;
- James Stronge's *Teacher Effectiveness Performance Evaluation System*; and
- The McREL's *Teacher Evaluation System* (ASCD, 2014).

The Shift in Teacher Evaluation. Linda Darling-Hammond – arguably the preeminent voice of teacher evaluation in the United States – claims that the role of teacher evaluation is being reshaped as a result of an increased focus on teachers in the policy environment, of the growing sophistication of basic and applied research on teaching and teacher evaluation, and of

willingness of practitioners to engage many of the difficult issues that evaluation poses (Millman & Darling-Hammond, 1991). Over time, federal, state and local policy has evolved to reflect this changing perspective of teacher evaluation and its subsequent effect on student performance. Beginning in the 20th century, federal, state and local policy evolved to reflect this changing perspective of teacher evaluation and its subsequent effect on student performance. Most present policies in the latter half of the 21st century champion the use of teacher evaluation systems as the solution to improving teaching quality and ultimately to addressing equity issues (Minnici, 2014).

Below is a timeline of relevant educator effectiveness policy with respect to teacher evaluation.

Figure 1
Timeline of Relevant Educator Effectiveness Policy (with respect to Teacher Evaluation)

1965	1981	1994	1998	2002	2004	2007	2008	2010	2012	2013	2015
Elementary and Secondary Education Act (ESEA)	Education and Consolidation Act (ECIA)	Improving America's School's Act (IASA)	Higher Education Act (HEA)	Reauthorization of ESEA (NCLB Act of 2002)	Reauthorization of IDEA	<ul style="list-style-type: none"> ▪ Teacher Incentive Fund (TIF) ▪ Higher Education Act (HEA) <i>Title II</i> 	<ul style="list-style-type: none"> ▪ America Recovery and Restoration Act (ARRA) ▪ Race to the Top (RTTT) ▪ Investing in Innovation (I₃) ▪ School Improvement Grant (SIG) 	<ul style="list-style-type: none"> ▪ Teacher Incentive Fund (TIF) Cohort 3 ▪ The Maryland Education Act of 2010 	<ul style="list-style-type: none"> ▪ ESEA Flexibility Waivers ▪ District RTTT ▪ TIF Cohort 4 	<ul style="list-style-type: none"> ▪ Forty-four states have updated educator evaluation and legislation rules ▪ Two states plan to implement educator evaluation as part of their ESEA Waiver (KS; ND) ▪ Two other states (MS; NM) have recently proposed legislation on teacher evaluation 	The Every Student Shall Succeed Act (ESSA)

Retrieved from <http://www.gtlcenter.org/tools-publications/educator-effectiveness-timeline>

In 1965, The Elementary and Secondary Education Act (ESEA) was part of President Lyndon B. Johnson’s Great Society program. Passed in 1965, ESEA created a clear role for the federal government in K-12 policy, offering more than \$1 billion a year in aid under its first statutory section, known as Title I, to districts to help cover the cost of educating disadvantaged students. The law has been reauthorized and changed more than half a dozen times since that initial legislation. And, for the most part, each new iteration has sought to expand the federal role in education. (Klein, 2015)

The No Child Left Behind (NCLB) law—the 2002 update of the Elementary and Secondary Education Act (ESEA) of 1965—effectively scaled up the federal role in holding schools accountable for student outcomes. NCLB was the product of a collaboration between civil rights and business groups, as well as both Democrats and Republicans on Capitol Hill and the Bush administration, which sought to advance American competitiveness and close the achievement gap between poor and minority students and their more advantaged peers.

Since 2002, NCLB has had an outsized impact on teaching, learning, and school improvement—and become increasingly controversial with educators and the general public (Klein, 2015). In exchange for federal funds that go to states and, subsequently, districts based on a formula based in part of the number of students living in poverty, states agreed to put in place several protections and measures to ensure these funds are being used to support these students. The three major tenets of NCLB are the Annual Testing; Accountability; and School Improvement. The act also requires states to provide “highly qualified” teachers to all students. Each state sets its own standards for what counts as “highly qualified,” though these standards did not have to take into account student academic performance or outcomes (Education Post, 2014).

In 2008, the Obama administration implemented a \$4.5 billion federal grant program, Race to the Top (RTTT), which set in motion a host of state and local policies, requiring educators to develop and implement rigorous teacher evaluation systems that assess teacher effectiveness using student learning as at least one of the multiple measures. The priority of this legislation was to ensure that great teachers and leaders exist in all school systems by casting light on the wide variation in teacher effectiveness within and between schools (Coggshall, Rasmussen, Colton, Milton, & Jacques, 2012). Lost in the clamor generated by these policies is the equal weight that Race to the Top developers placed on requiring grantees to use evaluation to inform decisions regarding “developing teachers and principals, including by providing relevant coaching, induction, and/or professional development” (U.S. Department of Education, 2010, Sec. D[2]iv[a]) as well as other personnel decisions. Moreover, it required that winning states ensure that participating districts “conduct annual evaluations of teachers and principals that include timely and constructive feedback [and] as

part of such evaluation provide teachers and principals with data on student growth for their students, classes, and schools” (U.S. Department of Education, 2010, Sec. D[2]iii).

Teacher Evaluation in Maryland. On August 24, 2010, Maryland was awarded one of the federal government’s coveted Race to the Top grants in the amount of \$250 million over four years (Motel, 2010). The Maryland State Department of Education (MSDE) worked with a variety of education organizations to develop a new evaluation system to improve professional development and growth opportunities for both teachers and principals. Incidentally, one of the four pillars in Maryland’s RTTT grant application consisted of *Improving Teachers and Leaders*. The retooled evaluation system accounted for both *professional practice* (50% of a teacher evaluation) such as planning and preparation, classroom environment, instruction, and professional responsibilities and *student achievement growth* (50%) (USDOE, 2014).

Like other states around the nation, Maryland attempted to meet the expectations of RTTT via a wide range of policy approaches, specifically, making changes to the rules for how evaluations would be conducted in the hopes of boosting student achievement among student subgroups. Basically, the prevailing thought was that a revamped teacher evaluation tool would yield better teachers which, in turn, would produce higher achieving students. Some of the most common RTTT policy themes as it pertained to teacher evaluation in assorted state laws across the country consisted of the following:

- Multiple observations
- Varying frequency of observation by teacher experience and performance
- Advance warning for classroom visits
- Examining student work
- Feedback on observations

Per the adoption of the COMAR regulation applying to the Evaluation of Teachers and Principals (i.e., 13a.07.09 – see Appendix A) as a result of the Education Reform Act of 2010, the State Board must establish standards for performance evaluation for both teachers and principals which include model performance evaluation criteria. As a result, such regulations would meet the requirements of the Elementary and Secondary Education Act (ESEA) Flexibility Request requirements in 2012. Specifically, COMAR 13a.07.09 requires annual evaluation of probationary teachers and those who have been rated “ineffective.” Teachers rated “effective” or “highly effective” are evaluated on a three-year cycle in which a teacher in year two or three of the cycle may be evaluated using the previous year’s professional practice rating and the most recent available data on student growth. Moreover, under COMAR regulations, student growth must be a significant factor in each educator’s evaluation.

Under the framework proposed by the Council on Educator Effectiveness, 50 percent of a teacher’s evaluation is based on qualitative measures (planning and preparation, instruction, classroom environment, and professional responsibilities), and 50 percent is based on student growth (20 percent local growth measures and 30 percent state growth measures). No single criterion may account for more than 35 percent of a teacher’s evaluation. The Council for Educator Effectiveness had proposed a decision rule that would require a teacher or principal to be at least effective in the student growth component to receive an overall rating of effective or highly-effective, but this is not yet finalized in regulations and may be changed (Mead, 2012).

The U.S. Department of Education had invited each State education agency (SEA) to request flexibility regarding specific requirements of the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the No Child Left Behind Act of 2002 (NCLB) in

exchange for rigorous and comprehensive State-developed plans designed to improve educational outcomes for all students, close achievement gaps, increase equity, and improve the quality of instruction. As a result of the Elementary and Secondary Education Act (ESEA) Flexibility Waivers, most states (45 states, the District of Columbia, Puerto Rico and the Bureau of Indian Education submitted requests for ESEA flexibility) have adopted a teacher evaluation system that would include growth in student achievement. Additionally, states also endorse using multiple measures of teacher performance when evaluating teachers (US Department of Education, 2016).

In December 2015, Congress passed the Every Student Succeeds Act (ESSA) to replace NCLB. ESSA - also known as the ESEA Reauthorization - moved in the opposite direction (as NCLB) as it seeks to pare back the federal role in K-12 education. Because of the impossibly high aim of NCLB (such that all students would be deemed “proficient” by 2014), paved the way for an era of extraordinary leverage for the U.S. Department of Education. During the Obama administration, states were granted flexibility from the law’s original proficiency goals, but only if they agreed to adopt policies – including teacher evaluation – that replaced test-based accountability with measures designed to improve outcomes from the inside out (Silverman, 2017).

In essence, the ESSA has ended the federal government’s involvement in prescribing and influencing teacher evaluation systems across the nation. ESSA does not require states to set up teacher evaluation systems based in “significant” part on students’ test scores, which was a key component of the U.S. Department of Education state-ESEA Flexibility Waiver system. The law permits states to re-design and submit descriptions of their new accountability systems to the U.S. Department of Education and states are highly likely to dial

back stringent teacher evaluation requirements in favor of flexibility for most districts (NEA, 2016). Various modalities of feedback on school culture, coincidentally, is in fact the one of the primary goals of ESSA and it will be interesting to see how schools adopt and implement this expectation moving forward.

Current criticisms of the Teacher Evaluation Process. In a 2012 interview, teacher evaluation guru Ms. Charlotte Danielson actually commented that “I’m deeply troubled by the transformation of teaching from a complex profession requiring nuanced judgment to the performance of certain behaviors that can be ticked off a checklist.” (Meyers, 2012). In fact, Danielson went on to say that “if all you do is judge teachers by test results, it (teacher evaluation) doesn’t tell you what you (teachers) should do differently.”(Bellafante, 2012). These statements underscore a shocking critique by Danielson of teacher evaluation by one of its chief architects who - twenty-one years ago - developed one of the most viable evaluation tools by which school systems could quantify teacher performance. Meyer (2012) went on to report during his interview with Danielson that she (Danielson) claimed that while “teaching is hard, assessing it (teaching) shouldn’t be.” Unfortunately, this is a common refrain uttered by most educators and helps to validate this notion that doing teacher evaluation accurately (and fairly) is easier said than done.

Practitioners, researchers, and policy makers agree that most current teacher evaluation systems do little to help teachers improve or to support personnel decision making. Given the fact that the focus of most conversation surrounding teacher evaluation is on its summative side, Danielson (2001) claims that such focus is displaced and detrimental to the education improvement process for teachers in terms of professional growth. There’s also a growing consensus that evidence of teacher contributions to student learning should be part of teacher

evaluation systems, along with evidence about the quality of teacher practices. (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). This stark reality may be conflating the issue of what defines “good teaching” with student test scores in the teacher evaluation process. While it seems reasonable for teachers to demonstrate that their kids have learned, it is far from certain that this process can be done fairly for teachers, particularly in a high stakes environment (Meyer, 2012). Regardless of the approach, states, districts, schools, principals and teachers have struggled with the shifts of waiver era teacher evaluation systems.

Federal policy goals, over time, have shifted from ensuring that all teachers had traditional credentials and were fully certified to creating incentives for states to evaluate and retain teachers based on their classroom performance (Whitehurst, Chingos, & Lindquist, 2015). Traditionally, even before there was a law mandating it, principals have long conducted teacher evaluations. Yet those traditional evaluations, typically based solely upon classroom observations, had little effect on teacher quality (Jacob & Lefgren, 2008). Teachers remained in place even if they were obviously struggling. And nearly every one of them got a satisfactory (or even “outstanding” rating). For instance, a California judge in the recently-decided *Vergara* case found that a significant number of “grossly incompetent” teachers were allowed to remain in the classroom “because school officials don’t want to go through the time and expense to investigate and prosecute” these cases (Peterson, 2014).

The controversy surrounding teacher evaluations has resulted in a moving target for educators. States and school systems across the country have been embroiled in conversations with numerous teachers unions as to how to accurately and reliably rate teachers as part of Race to the Top (Aldis, 2014). Second, whether fair or not, teacher evaluation systems are being portrayed publicly as being anti-teacher.

This is harmful in at least two ways: it has a negative effect on teacher morale, and the ensuing debate becomes polarized and focused on the wrong issue. As a result, teachers have developed an acute fear of teacher evaluation over time and the consequences that have arisen from those evaluations (Bogart, 2013). Frankly, the problem becomes compounded if the real problem is not defined (i.e., what components should go into a teacher evaluation model) and, summararily, the chance of finding a solution becomes elusive. Seemingly, states and schools systems are reaching a tipping point of bureaucratization and conversations about teacher evaluation systems are becoming increasingly divisive (Dougherty & Jacobs, 2015).

Currently, most teacher evaluations produce metrics of compliance that rarely inform or unpack nuances in teaching practice so that substantive changes can be made to yield the greatest potential for improving student performance (Jerald & Hook, 2011; Weisberg, Sexton, Mulhern, & Keeling, 2009). This result in large part can be attributed to the fact that most teacher evaluation systems are single-source, one size fits all summative instruments that cannot differentiate between novice and experienced teachers (Danielson & McGreal, 2000) used to evaluate multiple people, regardless of content, level, etc.

Despite support for replacing ineffective and seldom used satisfactory checklists with new evaluation systems that can provide timely, actionable information to improve performance, the challenges of the current approach to evaluation loom large. The lack of time for high-quality implementation, diminishing resources for principal training, and poor connection with professional learning resources raise questions about the utility of these new systems and the extent to which they will continue to be a lever for improvement under future presidential administrations (Silverman, 2017).

School leadership also posed challenges. Both evaluative capacity (the ability to conduct observations that accurately assess performance strengths and weaknesses) and relational trust (the extent to which principals developed an environment and individual trust relationships with teachers) affect the efficacy of an evaluation system (Silverman, 2017). Most recent research has addressed the ideal number of observations that should be conducted on an evaluatee (research suggests between two and three per year are recommended - Whitehurst, Chingos, & Lindquist, 2014). Furthermore, to mitigate any element of evaluator bias, some observations should be conducted by trained evaluators that are non-school based administrators that have limited substantial prior knowledge of the teacher being observed (Whitehurst, Chingos, & Lindquist, 2014; Ho & Kane, 2013; Kane & Staiger, 2012). The imperative of building evaluative capacity, coupled with the challenge of implementing evaluation programs, has left principals overwhelmed and underprepared (Silverman, 2017).

Additionally, the inordinately high ratio of evaluators to evaluatees preclude most stakeholders of taking the responsibility of using evaluations formatively (due to the overwhelming time constraints); instead, the model is built for meeting minimum requirements as infrequent meetings, inconsistent (and inconvenient) time frames for dialogue and less than timely feedback decrease reliability in the evaluation model with most parties feeling disenfranchised. Such specific instances fall under the umbrella of a “glaring lack of sufficient resources devoted to teacher evaluation” and the literature validate this phenomena nationwide (Stecher & Garet, 2014; Darling-Hammond Amrein-Beardsley Haertel & Rothstein 2012; Kane & Staiger, 2012; Jerald & Hook 2011; Goldhaber, 2010).

The top-down approach to reforming a long-broken component of school talent management is rife with problems. While many educators seemed to get behind improved

evaluation systems, few felt comfortable with the stakes. Including students' standardized test performance as a factor in teacher evaluation raised serious concerns. According to audits conducted in two states, only about 25 percent of courses taught in K-12 schools include a standardized assessment usable in a value-added metric. That leaves nearly three-quarters of educators with assessments of student learning that must be tracked and measured in a less standardized – and perhaps less reliable – way (Silverman, 2017).

Educators' reactions to evaluation systems varied consistently with the assessments considered in their evaluations. A 2016 report from the Institute for Education Sciences reported that teachers whose evaluations included student test scores were more than twice as likely to report dissatisfaction with their evaluation system. States and districts addressed these concerns by adopting alternative metrics for student outcomes, such as Student Learning Objective (SLO) systems, and reducing the role of assessment data within evaluation rubrics by including multiple years of outcomes or lowering their relative weight in the final evaluation rating (Silverman, 2017).

Improving the Teacher Evaluation Process. As we move further into the first decade of the 21st century, education stakeholders called for major changes in tenure and compensation as current evaluation practices sustained heavy criticisms, as the emphasis had shifted to evaluation from simply supervision. In 2005, Tucker and Stronge claimed that student achievement should be used as an important criterion in the evaluation process (Tucker & Stronge, 2005). Toch and Rothman purported in their 2008 *Rush to Judgment* report that the teaching profession focused on formal credentials rather than on instructional effectiveness and current supervisory and evaluative practices of teachers were “superficial, capricious, and often don't even directly address the quality of instruction, much less measure

students' learning" (p. 1).” To justify their point, only 14 states required school systems to do annual evaluations of teachers despite *No Child Left Behind* (NCLB) requirements about teacher quality (Toch and Rothman, 2008).

The acute challenges of teacher evaluation metrics to serve as either a quality indicator or as informative was illuminated in the study known as “*The Widget Effect*” (Weisberg, Sexton, Mulhern, & Keeling, 2009). The study examined the evaluation systems in 12 districts across four states and concluded that 99 percent of teachers received a satisfactory rating. The study described the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher. This decades-old fallacy fostered an environment in which teachers ceased to be understood as individual professionals, but rather as interchangeable parts. In its denial of individual strengths and weaknesses, it was deeply disrespectful to teachers; in its indifference to instructional effectiveness, it gambled with the lives of students (Weisberg, Sexton, Mulhern, & Keeling, 2009):

The Maryland State Department of Education (MSDE) – as directed by the Race to the Top (RTTT) Grant – has released teacher and principal evaluation ratings for the 2014 School Year statewide in the form of the Teacher Performance Evaluation School Detail File (MSDE Teacher and Principal Ratings Analysis and Data, 2014). The results are strikingly similar to The Widget Effect results in that teachers (and principals) were rated overwhelmingly as “effective” or “highly effective” with a fraction of all teachers/principals rated as “ineffective.” In fact, what was reported in the aforementioned Detail File across the entire State of Maryland, only 1.4% of all principals (i.e., 16 out of 1112) and 6.2% of all teacher (i.e., 2,149 out of 34,422) were rated as “ineffective.” Moreover, across some local districts, some schools systems had not one teacher or principal evaluated as “ineffective.”

Incidentally, the school system in this research study had zero principals (out of 26) and 4.96% of teachers (i.e., 27 out of 544 total teachers) rated as “ineffective” (MSDE Teacher and Principal Ratings Analysis and Data, 2014).

The Aspen Institute (2016) has noted the following experiences of both states and districts that highlight the key issues in the implementation of teacher evaluation systems in the new era of ESSA: (1) the proper division of responsibility between State Education Agencies and Local Education Agencies; (2) the training and capacity of principals and other classroom observers and evaluators; (3) the use of meaningful measures of student growth; (4) personalization of the evaluation and support process; (5) the tension between state formulas and professional judgement in assessing a broad body of evidence related to teacher performance and practice; (6) the degree of teachers’ involvement in decision-making about which aspects of their practice should be the focus of evaluations and in providing feedback on whether evaluations are helping improving practice.

States and districts are also likely to take a cue from new federal language elevating the importance of professional learning for teachers and principals. For the first time, states may set aside up to 3 percent of Title II(A) funding allocations to support principal training and development (Silverman, 2017). This shift signals an increased interest in the role of school leaders in developing and retaining the talent required to succeed with students. The new language also consolidates a specific definition of professional learning designed to mitigate low-quality training and drive up access to effective learning. Undoubtedly, as Silverman (2017) states: the key to professional learning is to connect it with teachers’ demonstrated needs – which means connecting it with data collected from observation and student outcomes. Taken together, states and districts will be wise to maintain evaluation

systems that provide actionable information and, ultimately, deliver on their original purpose: setting teachers (and students) up to succeed. One thing is certain: teachers' and principals' opinions will be a primary factor in reshaping teacher evaluation.

Effective teacher evaluation and professional growth. Some research suggests that teacher evaluations can make a difference in the effectiveness of teachers (Taylor & Tyler, 2012) and that the least skilled teachers benefit the most from thoughtful evaluations (Taylor & Tyler, 2011). What Taylor and Tyler have quantified is that mathematics teachers who had not been initially highly rated on their local evaluation system (i.e., Teacher Evaluation System- TES) in the Cincinnati Public Schools actually improved over time. The researchers attributed teacher improvement in the TES rating to the following components of the System: a detailed rubric describing practices shown to correlate positively with student mathematics achievement; multiple observations; feedback opportunities over the course of an entire school year, and regular evaluator training (Taylor & Tyler, 2011).

So the question becomes, what are the characteristics of an evaluation model that informs teachers about how to improve their practice while providing information to evaluators and other educational stakeholders (such as students and parents) as to which teachers are rated as the most (least) effective? As stated previously, as a result of the *Elementary and Secondary Education Act* (ESEA) Flexibility Waivers, most states have adopted a teacher evaluation system that would include growth in student achievement. Additionally, states also endorse using multiple measures of teacher performance when evaluating teachers. Various metrics for evaluative purposes should include student achievement data; classroom observations; student/parent surveys; lesson plan reviews; teacher self-assessments; measures of professional learning; student artifacts; and teacher

portfolios (Hull, 2013). Overwhelmingly, the literature endorses the utilization of multiple measures of teacher performance and, as a result, should be incorporated in evaluation models to serve the dual purpose of measurement and improvement in teacher quality (Ellett & Teddlie, 2003; Weisberg, Sexton, Mulhern & Keeling, 2009; Jerald & Hook, 2011; Darling-Hammond Amrein-Beardsley Haertel & Rothstein 2012).

Seemingly, for a teacher evaluation system to be both transparent and credible, teachers and administrators must understand what constitutes good practice; that is, do vested parties have an agreed-upon and consistent definition of good teaching and what behaviors constitute excellence in teaching? Additionally, are there ample opportunities to engage in meaningful conversations about teaching practice? No matter what model or metric with which teachers are rated, there must be an infrastructure in which conversations can happen so that school districts can enhance professional practices. And, finally, is there an acute focus on “what matters” in an observation; a focus on which teachers and administrators concentrate their collective attention on the salient issues of teaching and learning?

The challenge is merging the two overarching purposes of teacher evaluation: quality assurance with professional learning. Educators need to create procedures that yield valid and reliable results—that is, that satisfy the legitimate demands for quality assurance while promoting professional learning (Marzano, 2014). In truth, the demands are somewhat different. Marzano purports that for a system to ensure quality, there must be parameters that are "hard-sounding" qualities such as ‘valid, reliable, and defensible’ whereas a system designed to promote professional learning is likely to be collegial, introspective, and collaborative—which are much "softer-sounding" qualities.

Until recently, educators' attempts at merging quality assurance with professional learning have taken the form of enhancing evaluators' skills using techniques like clinical supervision and cognitive coaching. These are valuable skills and worth learning, but they are insufficient, claims Marzano. He claims that the profession is better served when the requirements for these two purposes are embedded in the design of the systems themselves (Marzano, 2014). The question at hand is are there viable strategies in which school system leaders can embed and promote these dual purposes into the existing teacher evaluation system?

Also to be considered is the assumption that various educator effectiveness policies across the nation (such as teacher evaluations) can actually yield improved teacher effectiveness. The research base on this assumption is not particularly deep. Yet, one of the strongest studies on this issue comes from a pair of economists, Eric Taylor and John Tyler, and has been cited previously in this dissertation. Taylor and Tyler examined the relationship between being evaluated and subsequent teacher value-added using data from a long-running evaluation system in Cincinnati. Certainly, solid, research-based and helpful evaluation and observation techniques of teachers can have a beneficial backwash on teaching. The issue is and remains: What kind of evaluations are being employed and to what ends? School leaders cannot use a checklist approach to observing teachers and providing feedback. Teacher observation requires a comprehensive model that acknowledges the segments that make up a lesson. As stated, a comprehensive observation method includes teachers' self-reflection, walkthroughs and formal observations by principals and peers.

From a "quality assurance" perspective, there must be a shared understanding of the definition by everyone in the system of what constitutes good teaching. Teachers, mentors, coaches, and supervisors— must possess common language to describe practice increases the

value of the conversations that ensue from classroom observations. For example, discussing "student engagement in learning" is more effective when everyone understands what this looks like in light of various elements such as activities and assignments, grouping of students, instructional materials and resources, and structure and pacing and, especially, in the context of the evaluation model that is being employed. Conversations using this more specific language invite teachers to analyze their own practice and invite observers to inquire about the decisions a teacher has made in planning and executing a lesson. Certainly, when the evaluator and the evaluatee jointly use any standards-based rubric associated with a teacher evaluation tool to help differentiate "effective" from "ineffective" demonstrable behaviors both the evaluators and the evaluatees are able to move an evaluation closer towards the goal of quality assurance.

Additionally, there is also an acute need for trained evaluators to exist so as to further ensure a credible system of teacher evaluation. A credible system of teacher evaluation requires higher levels of proficiency of evaluators than the old checklist, "drive-by" observation model. Evaluators need to be able to assess accurately, provide meaningful feedback, and engage teachers in productive conversations about practice. Those who support teachers—mentors, coaches, supervisors, and so on—must be able to recognize classroom examples of the different components of practice, interpret that evidence against specific levels of performance and engage teachers in productive conversations about their practice. Evaluators must be able to assess teachers accurately so teachers accept the judgments as valid and the public has confidence in the results. Evaluations that focus on quality assurance yield judgments that are fair, reliable, and valid. They are helpful in looking at both new and experienced teachers' practice and in determining whether a teacher's skill has slipped below

standard and needs strengthening. Administrators may use the evaluations for decisions regarding employment and compensation. This is crucial when deciding which teachers should attain permanent status as tenured professionals or which teachers should be nominated for leadership positions as mentors or coaches.

However, the existence of trained evaluators to be present in a school system may be easier said than done. The Danielson Group and their experience with the *Framework for Teaching* expressed mild frustration and even humility when most observers even after training require multiple opportunities to practice using the framework effectively and to calibrate their judgments with others (Danielson, 2011). Danielson claims that most administrator preparation programs don't teach such skills; administrators must acquire them on the job. But when they do learn them, administrators can be the instructional leaders that schools so urgently need (Danielson, 2011).

Danielson claims that a training program for evaluators—one that uses the *Framework for Teaching*—consists of the following four steps (see below); the persistent problem, however, is that new administrators are constantly being minted every year and –as a result – there seems to be a constant void of evaluators that are “adequately trained” annually. Danielson claims that administrators/trained evaluators should participate in the following so as to become a more seasoned evaluator in her teaching Framework:

- Participants familiarize themselves with the structure of the Framework for Teaching.
- Participants learn how to recognize the sources of evidence for each component and element.
- Participants learn how to interpret the evidence against the rubrics for each component's levels of performance.

- Participants learn how to calibrate their judgments against those of their colleagues.

A review of the literature finds a great deal of support for the premise that quality observations should be based on clear standards of teaching practice (Danielson & McGreal, 2000; Donaldson, 2009; Kane et al., 2010). Further, others have suggested that if the standards for evaluation are based on sound teaching practices and they are implemented in the classroom, student achievement will increase (Odden, 2004). Defining effective teaching and establishing domains of such standards are common within these systems. These standards of effective teaching would comprise the basis for performance measures to evaluate in the classroom. Further, these standards would define what practices need to be present in order to have an effective lesson (Milanowski & Heneman, 2001).

For some school systems that employ a standards-based teacher evaluation system, there is moderate consensus about exactly what “effective” and “highly effective” teaching look like for particular instructional skills (Danielson & McGreal, 2000). Jacob and McGovern in their report addressing teacher development, *The Mirage* (2015), cite the following example:

In order to earn a rating of “effective” in competencies aligned with developing students’ critical thinking skills, for example, teachers need to demonstrate to their observers that they are posing meaningful questions to students, which lead students to critically assess information and rely on evidence to put forth a point of view. To earn a “highly effective” rating in this same category, teachers must masterfully do so in such a way that all students lead their own conversations and are posing questions to each other. To earn a rating of “effective” at engaging students in lessons, teachers must be able to acknowledge student abilities and create

opportunities in response that result in most students being motivated by and equally engaged in appropriately challenging learning tasks. Those rated “highly effective” are able to do the same but for all students, leaving no one behind. These are complex skills, to be sure (p.17).

Danielson purports that her Framework for Teaching is an “evidence-based evaluation” and to earn these ratings, assorted demonstrable teacher behaviors are laid out in the Danielson Framework Rubric (see Appendix A for the Charlotte Danielson’s Framework modified for use in TPAS) so that the observer and teacher can – theoretically – jointly review the classroom evidence and, subsequently, discuss the teacher rating. Achieving “effective” instructional practice is not easy, and achieving “highly effective” practice is that much more challenging. But if we are going to get the results we need for students, teachers need to master these essential skills, and – subsequently - observers must assess teacher development efforts by how well they help teachers get there (Jacob & McGovern, 2015). As school systems attempt to merge the quality assurance piece of creating procedures that yield valid and reliable results of teacher evaluation with professional learning, one of the biggest and most formidable challenges a school system faces is finding the requisite time for professional conversations to promote professional learning. How do administrators and teachers find time to conduct meaningful observations and engage in professional conversations about practice? Evaluator-teacher conversations, when conducted around a common understanding of good teaching—and around evidence of that teaching— offer a rich opportunity for professional dialogue and growth. Since school systems cannot create more hours in the day, administrators and central office employees must build an infrastructure that embodies the careful setting of priorities and judicious scheduling of both observations and conferences so

that everyone can make the best use of the time available. Moreover, unless a district's negotiated agreement forbids it, brief and informal drop-in observations yield plenty of information for reflective conversation and require far less time than formal observations do (Roy & Heflebower, 2012).

In an *Education Week* article, James Stigler (2010) writes about the “lesson study” process in Japan, where teachers covering the same content meet regularly, develop their methods of student evaluation, and then meet together to examine the results. Stigler (2010) contrasts that system of teacher accountability with those being suggested by the Gates Foundation, Arne Duncan (Secretary of Education), etc. He says W. Edwards Deming would call what Gates and Duncan want “the inspection method.” In reality, Deming says, “real and continuous improvement occurs only when the workers themselves study outcome variability and the processes that produce it.” Deming’s alternative model begins with a well-defined goal and agreed-on measures for charting progress toward the goal. Soon thereafter, all workers should be involved in studying and improving the process that leads to the desired outcome, using “PDSA” cycles: First, *plan* an innovation, something worth trying; next, *do* it; then *study* the result of the change; and finally, *act*, whether by trying something else if the innovation didn’t work, modifying the innovation and going through the cycle again, or implementing the innovation as a permanent change in the production process (Stigler, 2010). Stigler’s article can be reviewed in conjunction with a study that has come out of the University of Chicago which reinforces the potential effectiveness of using trained teachers to give feedback to colleagues called Peer Assistance and Review (Thessin & Starr, 2011). In the study, teachers were far more demanding than principals using the same evaluation

system. The bottom line goal, nevertheless, is for feedback of *all* kinds to be part of the culture of the school.

Certainly, there is abundant evidence from both informal observation and formal investigation that indicates that a thoughtful approach to teacher evaluation—one that engages teachers in reflection and self-assessment—yields benefits far beyond the important goal of quality assurance (Goldstein & Noguera, 2006). Such an approach provides the vehicle for teacher growth and development by providing opportunities for professional conversation around agreed-on standards of practice and multiple measures of teacher performance (not just performance-based student outcomes) such as written work, portfolios, teacher-designed assessments. In essence, a comprehensive observation method should be the goal, which would include teachers' self-reflection, walk-throughs and formal observations by principals and peers, and instructional rounds led by teachers that acknowledges the segments that make up a lesson (Marzano, 2013).

As previously stated, Marzano claims that it is not possible to walk into a classroom and get a full perspective from a little slice of teaching unless you know exactly what is happening and which segment is occurring. That level of isolation, he says, has to change for teaching to improve (Marzano, Frontier, & Livingston, 2014). Marzano insists that administrators must supervise the art and science of teaching in a way that keeps attention focused on student learning rather than on specific instructional strategies. Unfortunately, what comprises most teacher evaluation system metrics nowadays are these very instructional strategies - so narrow in scope – that work against an individual teacher's flexibility and creativity in their own unique teaching practices.

The practical question that arises from such conclusions derived from teacher evaluations, though, is how can we facilitate real improvement in teacher skills in a way that can be afforded on a long-term basis by schools and school systems? Perhaps this responsibility should fall more on the evaluatee as opposed to the teacher evaluation system and/or the evaluator. Frankly, the power of any comprehensive teacher evaluation metric lies not simply in identifying the few ineffective teachers to fire or the few highly effective teachers to reward with increased compensation; rather, its primary role should be to facilitate the improvement of all teachers and establishing a framework in which a teacher evaluation metric can maximize the chances that those who remain in the profession are eventually rated as “highly effective” (and thus, considered accomplished practitioners). By focusing teacher evaluation on the continuous improvement of all teachers – not just for high stakes personnel decisions – such focus will have a greater impact on building teaching capacity in staff and, ultimately, improve student achievement.

So what should an effective teacher evaluation system look like? Jim Hull, a senior policy analyst for the Center for Public Education, wrote a report in 2013 entitled *Trends in Teacher Evaluation* (Hull, 2013) that stated the following five elements should be evident in a good teacher evaluation system: (1) Inclusive design and implementation process including teacher engagement; (2) Policies on how information will be used; (3) Multiple measures of effectiveness including: Data linking teachers and student achievement; (4) Classroom observations; and (5) Adequate resources and support.

Embedded in Hull’s aforementioned report was an examination of what different approaches varying states had incorporated into their systems in the hopes that education leaders could learn from each other to refine and improve their own metrics of teacher

evaluation. Ostensibly, Hull insisted that linking teacher performance to student achievement should be part and parcel of the evaluation since a teacher's main objective is to grow student learning. Moreover, because research has shown that observations are most effective when conducted (formally or informally) multiple times per year, school districts should consider training teams of observers since most administrators are ultra-busy and observations are more accurate when they are conducted by more than one person. Lastly, it was noted in the report that classroom observations are not the only measure of teacher practice. In fact, states and districts were found to rely on other forms of evidence to get a fuller picture of a teacher's performance and provide a rich, qualitative report about classroom instruction. More importantly, the components below add to the body of feedback teachers receive on their individual strengths and weaknesses and allow teacher practitioners to use such feedback to drive their formative growth as professionals, unique unto themselves. These include the following:

- Teacher self-reflection and self-assessment
- Lesson plan reviews
- Measures of professional learning
- Student artifacts
- Teacher portfolios
- Student/parent surveys

Other researchers like Darling-Hammond (2013) purport that criteria for an Effective Teacher Evaluation System should include the following 7 tenets: (1) Teacher evaluation should be based on professional teaching standards; (2) Evaluations should include multi-faceted evidence of teaching practice, student learning, and professional contributions; (3) Evaluators

should be knowledgeable about instruction and well-trained in the evaluation system; (4) Evaluation should be accompanied by useful feedback, and connected to professional development opportunities; (5) The evaluation system should value and encourage teacher collaboration; (6) Expert teachers should be part of the assistance and review process; and (7) Panels of teachers and administrators should oversee the evaluation process (Criteria for an Effective Teacher Evaluation System, 2014).

One of the more recent commentaries on teacher evaluation was published by The Aspen Institute in March of 2016 and written to help define a path towards improvement in this area. The position paper made ten specific recommendations for school systems to employ in the hopes of improving teacher practice and, subsequently, student achievement via an evaluation tool for teachers. The prevailing thought from decades of experience and research was that if school systems help teachers become better teachers, then students will achieve more in the classroom since teachers matter more to student achievement than any other aspect of schooling (*Teacher Evaluation and Support Systems: A Roadmap for Improvement*, 2016). The ten recommendations are as follows: (1) Prioritize principal and evaluator training and certification with a focus on professional growth; (2) Differentiate evaluation and support based on teachers' experience and past performance; (3) Allow teachers and observers to collaborate on areas of focus; (4) Allow for local discretion in accounting for student learning; (5) Respect the limitations of value-added data; (6) Support locally developed measures while pursuing improvements in their creation and use; (7) Make sure all important aspects of teaching performance are valued in evaluations; (8) Engage teachers in improving teacher evaluation systems; (9) Develop measures for testing the

integrity of evaluation system design and implementation; and, (10) Tell stories that go beyond performance ratings.

Self-Assessment in the National Board Certification Process. Many of the aforementioned “other evidence” reported by Hull (2013) and cited by states as components of their teacher evaluation system are already a part of the advanced teaching certification of the National Board for Professional Teaching Standards (NBPTS). The NBPTS certificate is achieved through a rigorous and voluntary assessment – not evaluation – program that candidates must take a series of assessments in content and pedagogy, including submitting videotapes as part of a comprehensive portfolio documenting their teaching skills. On average, the NBPTS process takes an estimated 200 to 400 hours to complete over the course of several years (Sawchuck, 2015). Interestingly enough, research has shown that students of National Board Certified Teachers (NBCT) have demonstrated an additional one to two months of instruction than their non-NBCT peers and the results of student performance are even more pronounced for minority and low-income students (Goldhaber and Anthony, 2007). Perhaps there are elements of the NBPTS process model that should be considered as viable addendums to not just the Danielson *Framework* and its derivatives but all teacher evaluation models that not only help teachers improve their own craft but, in the process, galvanize students themselves to improve their own performance.

Self-assessment is a powerful technique for improving performance and outlines a theory of teacher change. This theory provides avenues for peers and change agents to influence teacher practice that results in professional growth by teachers. Inherent in this theory is that teacher change occurs through reflection on experience and that self-efficacy beliefs mediate the influence of self-assessment on teacher practice (Ross and Bruce, 2007).

This theory of change is grounded in social cognition theory (Bandura, 1997) and by work derived from student self-assessment methodology (Ross, McDougall, & Hogaboam-Gray, 2002).

Fostering Reflection and Self-Assessment. Reflective thinking and self-assessment in teaching is associated with the work of Dewey (1933, 1938), who suggested that reflection begins with a *dilemma*. Dewey (1933) defines the process in reflective thinking as “an intellectualization of a problem to be solved.” Dewey sees reflective thinking as a strategy to address practical problems and equates it with the ‘problem-solving method’. Similarly, Lana M. Danielson (2009) extrapolates that such thinking (reflection) is embedded in the normal routine and repertoire of effective teachers. That is, highly-able teachers suspend making conclusions about a dilemma in order to gather information, study the problem, gain new knowledge, and come to a sound decision. This deliberate contemplation brings about new learning (L. Danielson, 2009). On a daily basis, teachers face a myriad of choices – dilemmas, if you will – on matters that toggle between the mundane and complex. Certainly, different types of thinking are needed to address such differentiated choices/dilemmas.

Seemingly, expert teachers adjust their thinking to accommodate the level of reflection for which a situation calls. Their teaching is characterized by an intentional competence that enables them to identify and replicate best practice, refine serendipitous practice, and avoid inferior practice. Because of their ability to reflect, great teachers know not only what to do, but also why; in essence, the very act of reflection helps to differentiate the “what” versus the “why” (L. Danielson, 2009). Research (Constantino & De Lorenzo, 2001; Danielson & McGreal, 2000; Glickman, 2002; Lambert, 2003) substantiates the role of reflection in teachers' professional growth. A disposition toward reflection—and a good sense of when the

teacher needs to step back and think deeply—should be part of all teachers' repertoires. The question ultimately becomes: How can school systems best nurture this habit of mind?

Lortie (1975) described how failing to reflect on teaching decisions leads to teaching by imitation rather than intentionality. Per Lana Danielson (2009), people who enter the profession have already gone through 16 years of "apprenticeship of observation" as students themselves and have developed preconceived ideas of what teaching is through having watched others do it. They may sense what teachers do but have no grasp of why they do it. Other researchers (Clift, Houston, & Pugach, 1990; Hargreaves & Fullan, 1992) have reinforced how important it is for teachers to examine their own beliefs about their classroom practices.

Danielson (2009) elaborates on Grimmett's *four modes of thinking* with regards to understanding the complexity of reflection: The four modes of thinking enable teachers to connect reflection to practical classroom applications. Danielson (2009) claims that when the following thinking modes are used appropriately, such *thinking* or *reflection* helps educators understand their own practice and, ultimately, foster the intentional competence necessary for accomplished teaching. *Technological or Formulaic Thinking* is defined as prepackaged thinking from an external source (L. Danielson, 2009) and works for many routine decisions (e.g., taking attendance, implementing emergency drills, etc.). *Situational Thinking* helps to shape decisions that are made on information embedded in a specific context at a specific moment (such as observing student behavior during a 2nd period class). This type of thinking does not look beyond the surface to consider root causes of problems (L. Danielson, 2009). *Deliberate Thinking* occurs when an educator purposefully seeks more information than the immediate context provides by revisiting theory, talking with colleagues, interviewing

students and/or student records. The goal with deliberate thinking is to better understand the dilemma (L. Danielson, 2009). The most pervasive type of thinking/reflection that one can do is *dialectical*. *Dialectical Thinking* builds on deliberate thinking to gain an understanding of a situation so as to generate solutions that are considered “new” and/or unconventional.

Basically, the greater a teacher’s ability to suspend judgment and the broader the repertoire of pedagogical strategies, the more flexible dialectical thinking will be (L. Danielson, 2009). In essence, this type of thinking categorizes a change in process and results in new teaching behaviors.

Succinctly put, teachers become confident about their future performance when they believe that through their own action they have helped children learn. A teacher self-assessment tool like the NBC model incorporates most of the following components: (1) influencing the teacher's definition of excellence in teaching and increasing his ability to recognize mastery experiences; (2) helping the teacher select improvement goals by providing him/her with clear standards of teaching, opportunities to find gaps between desired and actual practices, and a menu of options for action; (3) facilitating communication with the teacher's peer; and (4) increasing the influence of external change agents on teacher practice (Ross and Bruce, 2007). Ross and Bruce argue that providing a self-assessment tool (like the NBC process) is a constructive, professional development strategy for improving the effectiveness of teachers provided it is bundled with other professional growth strategies: peer coaching, observation by external change agents, and focused input on teaching strategies.

The NBPTS model of teacher evaluation is a recognized, research-based model for evaluating teaching practice that results in improved practice for teachers and performance by students. Established in 1987, the NBPTS represents one strategy for recognizing teacher

quality. The National Board is a voluntary system for assessing accomplished teaching. NBPTS offers an assessment process across several subject areas that is meant to signify teachers have achieved a high level of practice. NBPTS certification relies on an authentic, or “portfolio,” assessment process, which means that it uses artifacts of teacher practice, including videos of classroom lessons, student work, and reflective essays. Over the past two decades, both the program and the reach of National Board Certified Teachers (NBCTs) have grown substantially. Today, NBCTs number more than 100,000 and represent about 3 percent of the national teaching force (National Board of Professional Teaching Standards, 2010).

The portfolio is the cornerstone of the National Board Certification (NBC) process. On occasion, some candidates spend multiple years on the NBC journey as they reflect on their own practice and discover elements germane to their own teaching and learning that they would not otherwise have uncovered (Unrath, 2002). Teachers, in essence, are asked to create a picture of their professional life by videotaping themselves in action and, subsequently, commenting via written analyses of their video captures. As a result, candidates attempt to make the invisible and intangible aspects of accomplished teaching explicit by offering the teachers’ own planning, intentions, and analysis for the activity of lesson shown. These cases demonstrate the critical aspect of reflection – essential to conscious, ongoing improvement – including what the teacher recognize he or she could have done differently to be more effective (National Board for Professional Teaching Standards, 2016). Basically, as candidates go through the NBC process and build out their portfolio, teachers glean the unexpected benefit of careful deliberation; they metacognitively process their teaching choices to assemble a meaningful record of their lives as teachers and facilitates the ability of NBC candidates to think deeply about their professional practice as they conduct their own action

research (Unrath, 2002). This portfolio process of the NBPTS model is a penultimate example of a teacher's own unique way to "self-assess" as they aspire to improve.

Five Most Prominent Models of Teacher Evaluation in the United States. With respect to such recent and overwhelming policy focus in the United States on teacher evaluation, a number of evaluation tools have gained prominence, providing school leaders with a dizzying array of choices in terms of the models themselves, their accompanying implementation tools, and local system design decisions (ASCD, 2014). As previously stated, the RAND study in the 1980's concluded that teachers preferred specific as opposed to general feedback with regards to the actual practice of teacher evaluation in districts (Wise et al., 1984) and summarily laid the foundation in the late 20th century for the popularity of the Charlotte Danielson's Framework of Teaching (2000) to take hold. Additionally, school systems in the 21st century are beginning to see some evidence that the use of in-class observation data to drive meaningful conversations about instructional improvement can be valuable in both the desire to assess teacher practice and support professional growth (Darling-Hammond, 2009).

The Danielson Framework has been widely applied through K–12 as the evaluation model of choice for all teachers and continues to be one of the five most prominent teacher evaluation models of choice in school districts across the United States (ASCD, 2014). In addition to the Danielson Framework, the other most prevalent models are as follows:

- Kim Marshall's *Teacher Evaluation Rubric*;
- Robert Marzano's *Teacher Evaluation Model*;
- James Stronge's *Teacher Effectiveness Performance Evaluation System*; and
- The McREL's *Teacher Evaluation System* (ASCD, 2014).

No matter the teacher evaluation model that a school system employs, each has significant commonalities around *teaching expectations*. Therefore, embracing a robust approach to instructional design provides teachers and supervisors with guidance for success, whether one is the observer or the observed (ASCD, 2014). The researcher thought it would be a worthwhile exercise to unpack the aforementioned standards-based models and juxtapose the teaching expectation components of each; this is captured in the table below.

Figure 3

Components of the Five Models of Teacher Evaluation

Danielson	Marshall	Marzano	McREL	Stronge
Planning and Preparation	Planning and Preparation for Learning	Planning and Preparation	n/a	Instructional Planning
Instruction	Delivery of Instruction	Observable Classroom Strategies and Behaviors	Facilitate Learning	Instructional Delivery Standards
n/a	Monitoring Assessment and Follow-Up (includes a self-reflection form)	Reflection-on-Teaching	Reflection-on-Practice	Assessment of/for Learning (includes a self-assessment form)
Classroom Environment	Classroom Management	n/a	Respectful Environment	Learning Environment
Professional Responsibility	Professional Responsibility	Collegiality and Professionalism	Leadership	Professionalism
n/a	Family and Community Outreach	n/a	n/a	Student Progress
n/a	n/a	n/a	Know Content	Performance Standards

As we compare these five models, only two areas – sometimes referred to as “domains” in teacher evaluation vernacular – share a teacher expectation for which a teacher is to be evaluated throughout its model: *Instruction* and *Professional Responsibilities/Professionalism*. There does seem to be moderate consensus as to what is valued (and evaluated) amongst these systems: four out of the five models value *Planning and Preparation* (Danielson; Marshall; Marzano; and Stronge); *Learning Environment/Classroom Management* (Danielson; Marshall; McREL; and Stronge); and *Reflection* (Marshall; Marzano; McREL; and Stronge). Much less consensus was found across the models with respect to *Student Growth/Family Outreach* (Marshall and Stronge)

and *Content Knowledge* (McREL and Stronge). For the purposes of this study, what is most noticeable is the omission of reflection in the Danielson Framework.

Prior Attempts to Solve the Problem: The Teacher Performance Assessment System (TPAS). For close to two decades, the school district engaged in an extensive process of revisiting and revising the school system's process for the assessment of teaching. These revisions were designed to support the school system's commitment to an educational program that prepares students to develop their potential for a lifetime of learning and for responsible, productive participation in our diverse and changing world (Teacher Performance Assessment System, 2010).

In January 1998, the superintendent established a committee of exemplary teachers, union representatives, principals and central office leaders from the school system to develop teacher observation and evaluation instruments and guidelines for implementation. In her charge, the superintendent stated that, "it is essential that our teacher observation and evaluation process and tools reflect best practices in teaching and are consistent with the goals of the Maryland School Performance Program." She further stated that, "the key to improving instruction is improving the skill of the teacher."

The committee reviewed observation instruments and evaluation instruments from school districts in the state of Maryland and districts outside the state. They also reviewed certification by-laws, district policy, contract language, and the Professional Evaluation Program implemented in August 1990. At the end of January 1998, the committee was introduced to the work of Charlotte Danielson and her book, *Enhancing Professional Practice: A Framework for Teaching* (1996) (TPAS, 2010).

Throughout the spring of 1998, the committee reviewed documents and became familiar

with Danielson's work. An administrator who was Coordinator of Organization Development and Home Instruction for Howard County Public Schools, MD presented the work that Howard County had done to create an evaluation document based on Danielson's work. By summer 1998, the committee had adopted the work of Danielson and began the complex task of customizing the framework to meet the needs of the school system. During the 1998- 1999 school year, the district's committee gathered feedback from various stakeholders and made modifications.

During the 1999-2000 school year, the document was reviewed by a broader representation of the district. The on-going professional development for district administrators and supervisors intensified. Trainers from Educational Testing Service (ETS) joined the district in June 2000 to provide comprehensive professional development in the framework and in the use of the program adopted by the district to enhance the use of the *Framework* for all practitioners—principals, supervisors, and teachers.

In the 2000-2001 school year, each building administrator implemented the program with a minimum of two first year teachers. Professional development opportunities were provided to all first year teachers, their mentors, and the administrative and supervisory staff of the school system (TPAS, 2010).

In the 2001-2002 school year, all first and second year teachers began using the new system as well as staff members from the five pilot schools: one high school, one middle school and three elementary schools. Mentors for non-tenured teachers were trained in the program. In 2002-2003, the final component of the model, the formative assessment process, was implemented. All professional staff participated in the teacher performance assessment system (TPAS, 2010).

Over the past several years, the evaluation process was refined and revised; the intent was to ensure clarity, consistency, and fairness to all involved. The overarching objective of the process was to develop a mechanism by which teachers can receive feedback on the expectations for teaching and learning while constantly improving over the time employed in the school system.

Characteristics of TPAS. The current system of teacher evaluation in this district utilizes a schedule of pre-observation, observation, post-observation, and reflection that consume over four hours, with the evaluator (who must meet MSDE requirements for observation and evaluation). A summary of the district's Teacher Performance Assessment System (TPAS) reflects the following:

- The use of multiple sources of information to evaluate teaching in addition to direct classroom observation (for example student work, teacher artifacts, planning documents, teacher reflection);
- An emphasis on teacher self-assessment, reflection, and collegial support;
- A role for teacher autonomy in the assessment process, combined with adherence to accepted measurement principles in the assessment processes for accountability purposes; and
- The use of multi-year assessment cycles, with different procedures for teachers in different phases of the cycle.

The assessment system includes two different processes: a formative process, under the direction of the teacher, and a summative process, which involves administrators in making judgments regarding teaching performances. All probationary teachers are engaged in the summative process each year. Once teachers receive continuing contracts, however, they

participate annually in either the formative process or the summative process (TPAS, 2010). Specifically, all of the district's teachers are on one of the following three evaluation cycles:

- Summative Non-Tenured
- Summative Tenured
- Formative

If a teacher is on a "Summative" evaluation cycle (be it tenured or non-tenured), then these teachers are to be formally observed using TPAS four times in the school year. Each of the four observations would include a pre-conference and a post conference in addition to the actual observation. Contractually, teachers must have their first observation cycle completed by Winter Break and the second cycle finished by April 15 of the school year. Evaluators are to make every effort possible to schedule observations with the appropriate content supervisor so as to condense the four scheduled observations to two (since, in our county, if a school based administrators and supervisor conduct an observation together, it would count as "two" observations as opposed to the required "four."

Limitations of TPAS. Presently, the school district in the study uses an evaluation system (TPAS) which is a derivative of the Danielson Framework. The TPAS is a system that is ostensibly in-depth and rigorous. However, there are several problems with its implementation across school sites. For one, school systems that employ some derivative of the Danielson Framework, there is *moderate* consensus only about exactly what "effective" and "highly effective" teaching look like for particular instructional skills (Danielson & McGreal, 2000); an example from Jacob and McGovern's *The Mirage* (2015) was previously cited addressing this particular issue.

Danielson purports that her Framework for Teaching is an “evidence-based evaluation” and to earn these ratings, assorted demonstrable teacher behaviors are laid out in the Danielson Framework Rubric (see Appendix A for the Charlotte Danielson’s Framework modified for use in TPAS) so that the observer and teacher can – theoretically – jointly review the classroom evidence and, subsequently, discuss the teacher rating. Achieving “effective” instructional practice is not easy, and achieving “highly effective” practice is that much more challenging. But if we are going to get the results we need for students, teachers need to master these essential skills, and – subsequently - observers must assess teacher development efforts by how well they help teachers get there (Jacob & McGovern, 2015).

Certainly, the limitations that exist in the local TPAS model of evaluation are not dissimilar to the criticisms that other evaluation systems have experienced and the researcher has previously noted. These include insufficient time to provide the mentoring and coaching needed by some teachers to grow professionally. Moreover, like most teacher evaluation systems, TPAS tends to be labor intensive, sometimes to an overwhelming degree. Not to mention, the preponderance of inconsistency of implementation by observers and evaluators can also be a problem. Certainly, the lack of priority given to the teacher evaluation process by both teachers and evaluators can limit the potential of the evaluation process to advance professional growth. Despite the best intentions and efforts of all involved there is still a void in the formal TPAS process that is limiting its effectiveness. This researcher will study some of the factors that have the potential to fill the void in the teacher evaluation process. By strengthening the teacher’s role in the process and empowering teachers to take more overt responsibility for their own professional growth, it is assumed that the evaluation tool be used more effectively - i.e., formatively – to improve teaching practice.

Given the pause in the national landscape regarding teacher evaluation with the recent passage of the ESSA, states and locals should seriously consider this opportunity as a way to move and/or differentiate teacher evaluation systems – including the Danielson Teacher Evaluation Model and its corresponding derivative models - to a more formative framework for teachers using the reflective elements of the National Board Certification (NBCT) process. The thought being that a more formal incorporation of an element of the NBCT model which may not necessarily be found in a local school system’s evaluation tool such as teacher self-reflection (a paradigm of the NBCT model) could galvanize professional teacher growth, which in turn would yield higher rates of student performance. Let’s juxtapose the Danielson Framework against the NBCT model to see if there are salient differences between the two models.

Instrumentation Tool Differences	
Danielson Framework (TPAS and derivatives)	National Board (NBCT) Model
<ul style="list-style-type: none"> • Comprised of Four Domains • An internal <i>evaluation</i> of teacher performance • Involuntary • Rubric-based • A <i>snap-shot in time</i>, no more than 45 to 120 minutes of pre-conference, teacher observation, and post-conference • Logorrheic in nature but narrowed down to a related totaled number depending on the value assigned to a teacher within each level of the rubric • Teachers are expected to explicitly use data as part of the Framework • Teachers are expected to demonstrate professionalism regarding what a teacher should know and be able to do. 	<ul style="list-style-type: none"> • Comprised of Five Core Propositions • An external <i>assessment</i> of teacher self-assessment • Voluntary • Portfolio-based • Assessment based (content knowledge) • A <i>one year encapsulation</i> of teacher performance that combines content assessments and the submission of a portfolio to a peer-review committee for each certificate. • Rigorous, teacher assessment program that includes video-tapes and written responses/reflections of teachers • Developed to create tools to define and measure teacher excellence that is individualized • Home communication is documented and acknowledges that a changing family structure for students presents new challenges for teachers • Teacher-reflection is quantified (in 4 ways)
Differences between the Danielson Framework's <i>Four Domains</i> with NBCT's <i>Five Core Propositions</i>	
Danielson Framework (TPAS and derivatives)	National Board (NBCT) Model
<ul style="list-style-type: none"> • Absence of explicit mention of technology in the Framework • Clear standards of student conduct are assessed in the Framework • A focus on oral questioning and encouraging high levels of questioning in the Framework • Using data to guide instruction is a component of the Framework • Showing/demonstrating professionalism is a component of the Framework • Limited formal use of focused teacher reflection is evident in the Framework • Teacher assessment and/or an instrument to evaluate students is part of the Framework 	<ul style="list-style-type: none"> • Heuristic learning and/or self-discovery approach to learning • Limited to zero reference or adoption of any particular discipline model; not assessed in the model • No explicit reference to using data to drive instruction; not assessed in the model • Speaks more to student engagement and not specifically referencing oral questioning technique of the teacher • Not explicit reference to a teacher's assessment and/or an instrument to evaluate students; not assessed in the model • Professionalism is not a critical component for teachers to include in their portfolio; not assessed in the model

Figure 3: The Differences between the Danielson Framework and the National Board Model

Given the table above, there are certainly differences between the two frameworks. Primarily, the Danielson Framework (and its derivatives) is more of an internal *evaluation* (rubric-based) of an ephemeral teacher performance performed by evaluators as compared to the NBCT model that is a voluntarily-driven external *assessment* of a teacher, captured in a portfolio, which is reflective of a one's own self-efficacy, reflection and assessment over a year's time.

Perhaps the issue boils down to the subtlety between a teacher performance system that is categorized more as an *evaluation* (TPAS/Danielson) as opposed to an assessment (NBCT). The Merriam Webster textbook definition of evaluate means to determine the significance, worth, or condition of, usually by careful appraisal and study." Assess means "to determine the importance, size, of value of" (Merriam Webster, n.d.) as in assess the problem or assess if there is a problem. As Vivano purports, when an administrator evaluates a teacher, he or she is placing worth on another human being's skills. Contrast that with when an administrator assesses or a teacher self-assesses, they are looking for the magnitude of a problem or even if a problem exists. If a problem or lack of teacher skill in a certain area is noted, one can then make improvements based on the actual assessment (as opposed to the evaluation) of said teacher performance. After assessment, the administrator and the teacher jointly can then summarily concern themselves with what can be done to ameliorate any problems or skill deficiencies that were revealed during the assessment process for the sole purpose of helping the teacher improve their practice. Much like what teachers do on a daily basis with their own students, administrators could work in concert with these teachers and assume the role of facilitator to make sure that the teacher goes through the appropriate steps and professional development to improve their practice based on the aforementioned assessment of a teacher's performance (Vivano, 2012).

For most school systems, the Danielson Framework (and its derivatives) – in its present deployment – is somewhat limited in its scope of being utilized *formatively* as its chief aim is inherently dependent on how local school systems choose to use said Framework. The researcher would like to know if there is evidence that teacher quality improves over time (in the form of higher teacher ratings) as teachers are evaluated over that *same* time. Nevertheless, as Charlotte Danielson herself states, the intention of her widely used Framework was primarily to be utilized for teacher “self-assessment” for purposes of improving teacher practice:

When I wrote *Enhancing Professional Practice* in 1996, I intended it to be a definition of good teaching, in all its complexity. I hoped (and wrote) that it might be useful for any number of purposes: first, and most importantly, for teachers’ own self-assessment and reflection; for teacher preparation, recruitment and hiring, mentoring and induction; for professional development; and yes, also teacher evaluation. The latter was simply one of many uses to which it could be put.

Summary of literature review

In the past, teacher evaluation has generally not been a high stakes activity in part because improving the quality of teachers has not been seen as critical for improving the quality of education. Instead, school improvement efforts over the past several decades have focused on improving the curriculum, altering school management methods, and developing new programs (Millman, & Darling-Hammond, 1991). Thus, teacher evaluation, where practiced, was often an exercise to which few resources and little organizational attention were devoted. As a result, teacher evaluation has historically had little influence on decisions about personnel, staff development, or the structure of teaching. As times change, teacher evaluation results and research thereof is being used for a greater range of decisions and its

role in shaping teaching, as a whole will increase. Educators must consider how evaluation affects teacher performance, rather than whether or not evaluation will affect it (Darling-Hammond, 2000)

Presently, there is an unprecedented desire to improve the quality of information about teaching effectiveness via a credible teacher evaluation model. The hope is that the plethora of quantitative and qualitative data available to education professionals will help to guide local education agencies to build fair and reliable systems for measuring teacher effectiveness. But, as Minnici ponders, “*Can teacher evaluation systems, as currently designed and implemented, improve teaching practices?*” (Minnici, 2014).

There is a definitive need to build an evaluation system that gives trustworthy and reliable feedback to all stakeholders about teacher effectiveness and how to improve one’s own practice that directly impacts outcomes for students; especially when teacher effectiveness ratings are tied to quantifiable student achievement. In fact, when school systems compile datasets of student performance and integrate such data strategically into teacher evaluations, careful statistical analyses of these new datasets confirm the long-held intuition of most teachers, students, and parents: that teachers vary substantially in their ability to promote student achievement growth (Kane, Taylor, Tyler, & Wooten, 2011). As a result, it would make sense for evaluation systems to incorporate student performance data that make it possible to track the achievement of individual students from one year to the next, and to compare the progress made by similar students assigned to different teachers.

Metrics for evaluative purposes should include student achievement data; classroom observations; student/parent surveys; lesson plan reviews; teacher self-assessments; measures of professional learning; student artifacts; and teacher portfolios (Hull, 2013). Certainly,

when the evaluator and the evaluatee jointly use the TPAS rubric to help differentiate “effective” from “ineffective” demonstrable behaviors, both the evaluators and the evaluatees are able to move an evaluation closer towards the goal of quality assurance.

As school systems attempt to merge the quality assurance piece of creating procedures that yield valid and reliable results of teacher evaluation with professional learning, one of the biggest and most formidable challenges a school system faces is finding the requisite time for professional conversations to promote professional learning. How do administrators and teachers find time to conduct meaningful observations and engage in professional conversations about practice? Evaluator-teacher conversations, when conducted around a common understanding of good teaching—and around evidence of that teaching—offer a rich opportunity for professional dialogue and growth. Since school systems cannot create more hours in the day, administrators and central office employees must build an infrastructure that embodies the careful setting of priorities and judicious scheduling of both observations and conferences so that everyone can make the best use of the time available. Moreover, unless a district's negotiated agreement forbids it, brief and informal drop-in observations yield plenty of information for reflective conversation and require far less time than formal observations do (Danielson, 2011).

Certainly, there is abundant evidence from both informal observation and formal investigation that indicates that a thoughtful approach to teacher evaluation—one that engages teachers in reflection and self-assessment—yields benefits far beyond the important goal of quality assurance (Goldstein & Noguera, 2006). Such an approach provides the vehicle for teacher growth and development by providing opportunities for professional conversation around agreed-on standards of practice and multiple measures of teacher performance (not just

performance-based student outcomes) such as written work, portfolios, teacher-designed assessments. In essence, a comprehensive observation method should be the goal, which would include teachers' self-reflection, walk throughs and formal observations by principals and peers, and instructional rounds led by teachers that acknowledge the segments that make up a lesson.

Proposed Investigation

The overarching purpose of this research triangulates teacher evaluation, self-reflection and their roles in improving teacher quality. The prevailing thought is that teachers who willingly engage in more formalized self-reflection and self-assessment yield higher degrees of teacher effectiveness as measured on a local teacher evaluation framework known as the Teacher Performance Assessment System (TPAS). The central focus of this study investigated teachers' perceptions of the effect of TPAS on teacher quality – drilled down to each of the four component domains of TPAS - and other factors that contribute to a teacher's improvement of instructional performance over time. The researcher also reviewed the extent to which teacher cohorts – differentiated by demographic data - engaged in formalized practices of self-reflection about their own teaching practice. Lastly, the researcher also determines whether or not tenured teachers who are evaluated with TPAS actually improve their teacher effectiveness over time (as measured by TPAS).

SECTION II: METHODOLOGY

Purpose of the study

“Can teacher evaluation systems, as currently designed and implemented, improve teaching practices?” (Minnici, 2014). Currently, there is an unprecedented desire to improve the quality of information about teaching effectiveness via a credible teacher evaluation metric (Darling-Hammond, 2000). In practice, however, most evaluation systems have been generally used as sorting mechanisms for identifying the lowest performing teachers for selective terminations and meeting minimal compliance obligations while providing little incentive for teachers to improve their instructional practices. It is generally accepted that there is a two-fold need to build an evaluation system that not only provides reliable feedback to all stakeholders about teacher effectiveness but also informs teachers about how to improve their teaching practice. The challenge is merging the two overarching purposes of teacher evaluation: quality assurance with professional learning. Ideal components for evaluative purposes should include most, if not all, of the following: student achievement data; classroom observations; student/parent surveys; lesson plan reviews; teacher self-assessments; measures of professional learning; student artifacts; and teacher portfolios (Hull, 2013).

The central focus of this study was to investigate teachers’ perceptions of the effect that the Teacher Performance Assessment System (TPAS) has on teacher quality – drilled down to each of the four component domains of TPAS - and other factors that contribute to a teacher’s improvement of instructional performance over time.

The researcher wanted to investigate the extent to which teacher cohorts – differentiated by demographic data such as varying experiences, background, education, and/or training - engaged in formalized practices of self-reflection about their own teaching

practice. This study also investigated various cohorts' perceived degrees of engagement and willingness to participate in more formalized practices of self-reflection; ultimately, examining its collective impact on teacher quality in a local school system.

One of the goals of this study is to determine which particular teacher cohorts (disaggregated by demographic data) are rated "highly effective" more often than other cohorts and if certain teacher cohorts, in general, believe that engaging in self-reflecting behaviors positively impact their own teaching practice. The accompanying hypothesis is that as teachers become more apt to "self-reflect" on their own practice, they are more likely to grow as professionals, yielding higher ratings in any teacher evaluation model and thus becoming an accomplished, "highly-effective" practitioner. This researcher hopes to validate the work of Goldstein and Noguera (2006) that those particular school systems (and individual teachers, for that matter) that construct a more thoughtful approach to teacher evaluation - one that engages teachers in reflection and self-assessment - yields benefits far beyond the important goal of quality assurance (Goldstein & Noguera, 2006). Such an approach provides the vehicle for teacher growth and development by providing opportunities for professional conversation around agreed-upon standards of practice and multiple measures of teacher performance such as written work, portfolios, and teacher-designed assessments.

The overarching purpose of this research was to consider teacher evaluation and its role in improving teacher quality. The prevailing thought is that teachers who willingly engage in more formalized self-reflection and self-assessment – similar to the practices in which teachers who strive for National Board Certification (NBC) participate - yield higher degrees of teacher effectiveness as measured on a local teacher evaluation framework known as the Teacher Performance Assessment System (TPAS). For the purposes of this study,

various acts of self-reflection are analogous to acts of self-assessment and all future references to either self-reflection and/or self-assessment should be deemed interchangeable.

The research was conducted through a web-based survey using Qualtrics. The analysis of the data collection using surveys was designed to inform the researcher about teacher perceptions of the current evaluation metric (TPAS), the willingness of disaggregated teacher cohorts to participate in self-reflection activities, quantify the frequency *and* impact of such reflection, and to document the factors that impact teacher quality over time from the vantage point of a teacher.

Research Questions

The following research questions form the basis of this study:

1. Do tenured teachers believe that the current teacher evaluation system used in their district – the Teacher Performance Assessment System (TPAS) - impacts the quality of their teaching?
2. Do tenured teachers believe that being evaluated in the Teacher Performance Assessment System (TPAS) and its four component domains improve teachers' effectiveness over time?
 - a. *Domain I: Planning and Preparation*
 - b. *Domain II: Instruction*
 - c. *Domain III: Classroom Environment*
 - d. *Domain IV: Professional Responsibility*
3. To what extent do tenured teachers engage in self-reflection about their teaching?

4. Do tenured teachers that have been evaluated in the Teacher Performance Assessment System (TPAS) grow in teacher effectiveness over time (as measured by TPAS ratings) and, if so, what factors account for this growth?

Research Design

This quantitative study surveyed tenured teachers in one school system and examined teacher perception of their local teacher evaluation framework and its corresponding impact on teacher effectiveness. Quantitative data was collected via a 29-question survey with an opportunity for the participant to add comments on a few of the survey questions. Surveys determined demographic information of the teachers; teachers self-reporting their first/last formal TPAS evaluation rating; declarations of engagement of self-reflection including the frequency and impact of such reflection; and teachers' opinions on what factors/professional development activities most (least) improved one's teacher effectiveness over time. Surveys are an effective and expedient method to measure perceptions at a given point in time (Creswell, 2005). Using surveys is a common vehicle to gather nonexperimental data about a population which permits comparison between two groups within a population (McMillan, 2004).

Specifically, a comparative, cross-sectional survey design was used to identify of whether or not significant differences exist amongst survey respondents of various teacher cohorts from such demographics such as years of experience, highest degree conferred, National Board Certified Teachers (NBCTs), non-NBCTs, and those teachers rated as "highly effective" on the local evaluation metric. A t-test allowed a comparative investigation of the responses from the aforementioned multiple cohorts. Because this study utilized quantitative methods, a concentration on objective measurements and the use of statistical analysis of the

data collected occurred via the survey. Johnson and Onwuegbuzie (2004) encourage a quantitative approach to research in order to gain an objective understanding of the issue at hand.

These surveys were designed to address the teacher evaluation metric (TPAS) around a variety of topics, ranging from teachers' reports of their participation in professional development activities to their mindsets around growth and development as a teacher. The teacher online survey asked teachers to self-report on their own professional development, personal experiences within the teacher evaluation model (TPAS), recollect their own personal TPAS ratings throughout their tenure as a school system employee, and to assess their confidence in TPAS' ability as a tool to support teacher development. The researcher was looking for trends in the online survey differentiated by a teacher's experiences, background, education, and/or training. Survey participants have self-reported their demographic information as well on this survey. Additionally, the online survey was designed to consider the likelihood that teachers have engaged in degrees of self-reflective behaviors (both during the TPAS process and as part of their normal teaching routine). Also reviewed were the mindsets of teachers and the possible existence of a differentiated willingness to participate in self-reflection due to one's demographic background.

The survey results were collated by experiences, background, education, and/or training in conjunction with the self-reported TPAS rating. Moreover, the results were also disaggregated by respective degrees to which each demographic cohort engaged and/or was willing to engage in formalized practices of self-reflection to improve their instruction. Subsequently, the researcher generated the mean TPAS rating for each differentiated demographic/experience such as NBCT, non-NBCT, etc. To examine what extent teachers of

varying experiences, backgrounds, education, and/or training were rated higher (lower) on a teacher performance metric (e.g., TPAS) than others, have more (less) confidence in TPAS, engage and/or were willing to engage in self-reflective behaviors (or not), this researcher used a combination of descriptive analyses, regression, correlation and, cross tabulational analysis, which is one of the data analytics methods embedded into the Qualtrics Research Suite.

A cross tabulation table (also known as a “contingency table”) basically captures the frequency distribution of multiple variables and their interrelations (if any). This approach was first described by Karl Pearson in 1904 (Hai-Jew, 2017). A computational cross tabulation afforded the research the ability to seamlessly enable the identification of patterns in the teacher survey question responses at computer speeds with large amounts of data.

Teacher-Level Analysis: To investigate potential differences between teachers who did and did not improve over time, performance data will be linked to survey data. First, the researcher performed simple descriptive analyses, cross tabulations, Chi-Square analyses, and t-tests to determine generic data trends will differ significantly in terms of the following:

- The type of professional learning experiences, backgrounds, and trainings
- The presence of certain teacher mindsets regarding self-reflection and
- The presence and dosage of teachers engaging in self-reflective behaviors.

Additionally, the researcher performed a further series of analyses to investigate potential relationships between teacher performance and increased teacher support efforts, increasingly positive teacher mindsets when it comes to self-reflection and teacher perceptions of their environment on performance; all with respect to TPAS. The researcher inspected all items in separate models, controlling for years of teaching experience, training, education, and prior performance. Moreover, the researcher sought to determine whether teachers who had

more “optimal” development experiences with TPAS would be expected to have higher performance by regressing the various survey constructs in combination with each other. The researcher has also perform a series of cross tabulational analyses, using the same set of survey constructs, to test whether or not certain development experiences such as National Board Certification, positive mindsets with regards to self-reflection, or environments that facilitated teachers to engage in self-reflective behaviors increased the likelihood of being identified as an *improver*.

Design and Methods

Pretesting the Instruments. Content validity is described by Gay, Mills, and Airasian (2011) as the degree to which a test measures what it is intended to evaluate and item validity as the accuracy of test items in measuring the intended outcome of the research questions. The online survey was piloted with various administrators, counselors, and tenured teachers ($n=1420$) in the school district so as to solicit feedback on the survey. All participants were sent a link to the Qualtrics survey and were asked to provide written feedback regarding the items listed on survey to ensure its validity and alignment to the research questions. All feedback was considered and appropriate revisions were made.

SMCPS’ Department of Assessment and Accountability disseminated an email communique to all teachers in the school system with an embedded link to the Qualtrics survey. The survey included the University of Maryland research consent form with the described right of the participant to voluntarily participate (or not) in the survey. The consent form (Appendix D) described the level of confidentiality that will be provided throughout the research study.

Instrument. Participants were provided an online Qualtrics survey that included requests for demographic information, selected response, and Likert-scale questions. The instrument took approximately 10 minutes to complete. The sections of the online survey were aligned with the themes of the research questions, perceptions of teachers' confidence in TPAS; the existence, frequency, and impact of self-reflection on one's own teaching practice; and the various ways in which teachers perceive that they improve their own practice.

Limitations of the Study. The sample size the researcher used was confined to one school system and much of the quantitative data collected via the online Qualtrics survey was dependent upon the veracity of a teacher to accurately self-report. Another limitation in doing this study was that the researcher also serves as the county's Director of Assessment and Accountability; some respondents may not have answered as straightforwardly as a result. Additionally, surveying teachers that report either directly to the researcher and/or to principals that interface with the research could also possess an inherent bias when they were asked to ascertain the effectiveness of the present teacher evaluation metric that is currently used in the classroom. To also mitigate a portion of this bias, the researcher's surveys were anonymous with no identifying information and coded through Qualtrics software, which allowed for participants to complete the survey without collecting IP addresses, names, or other identifying factors.

More globally, however, it should be noted that education is relational and therefore provides an opportunity for bias. As stated by Lunenberg and Irby (2008), a study's limitations are not under the control of the researcher. Teacher attitudes, level of professional development, years of experience, highest degree conferred, past experiences with the teacher

evaluation tool and/or the number of formal observations with which one has been subjected could contribute to a particular bias. Lastly, surveys do not provide information that allows cause and effect explanations of data (Creswell, 2005). Another disadvantage of surveys is the potential for a low response rate percentage which may result in sample bias in which some members or groups within the general population are under reported or unreported because volunteer respondents and non-respondents may differ in important ways (Creswell, 2005; Gay et al., 2009; McMillan, 2004).

Conceptual Framework

The researcher concentrated on four explicit areas that guided the research questions for this study regarding the improvement of teacher effectiveness over time. This conceptual framework assisted the researcher in the survey structure, instrumentation, research design, and the corresponding statistical analyses of the data collected. Figure 4 provides a visual representation of the conceptual framework which shows the factors that this study investigated in order to triangulate teacher evaluation, self-reflection, and their roles in improving teacher quality.

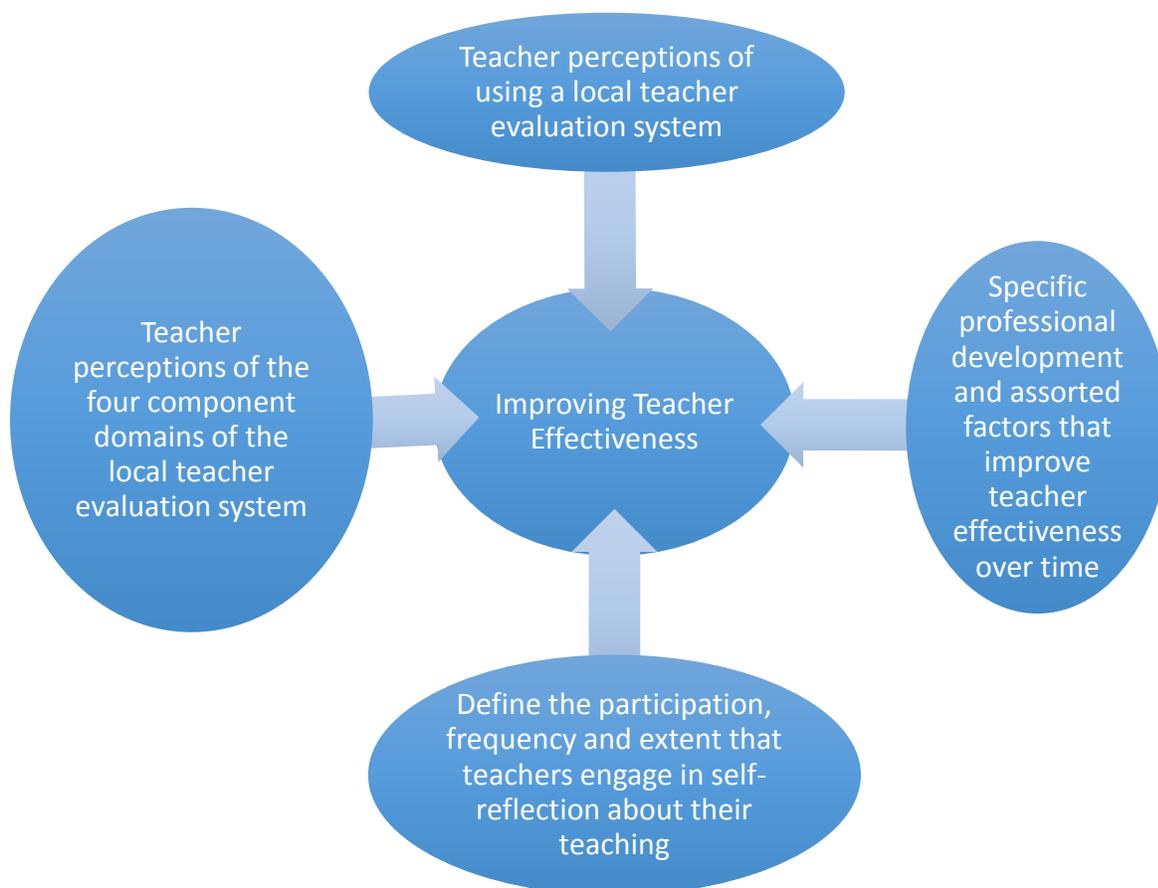


FIGURE 4. Conceptual Framework for Research Questions

Population and Sample. The school system used in this study is a rural school district in the southern Maryland region. Presently, the district is comprised of 18,067 students, 2,100 staff, 1,375 teachers, 121 administrators, and 29 school sites which includes 18 elementary schools, 4 middle schools, 3 comprehensive high school, 1 technical center, 1 K-8 school, and 2 intervention school sites.

The research was designed to examine the perceptions of all *tenured* teachers regarding the level of confidence in the current teacher evaluation systems (TPAS), the impact of self-reflection on one's own teacher practice, and the specific ways in which teachers improved their craft. Only tenured teachers had been included in the results of the study since said population was dually able to reflect on their experiences and daily routine as a teacher as well as having been rated by an administrator using the TPAS system.

- Originally developed in 1996, the Charlotte Danielson Framework for Teaching is used nationally to document and develop teaching practice. The school system in this research study deployed a modified version of the Danielson Framework named the Teacher Performance Assessment System (TPAS) used as a metric to evaluate teacher performance.
- All teachers in the school system are either on a “*summative*” or “*formative*” cycle for evaluation.
- As of 2014, there were 1420 teachers employed in the schools system with 13.87 average years of service. 95% of the school system's teacher are considered “highly qualified” per the Maryland State Department of Education (MSDE) and presently 39 teachers that have achieved National Board Certification.

- To mitigate the effect of a non-tenured teacher (since non-tenured teachers are not eligible to be considered for NBCT), the non-NBCT teacher cohort would only include tenured teachers in the school system. The research would be relegated to tenured teachers in the school system for the sole purpose of mitigating the effect of a non-tenured teacher (since non-tenured teachers are not eligible to be considered for NBCT); hence, the non-NBCT teacher cohort would only include tenured teachers in the school system.
- Classroom observation data consist of two types: *joint* and *solo*. *Joint observations* are defined as school leaders (principals and assistant principals at the school) and school system personnel (external observers) saw the same lesson at the same time but rated that lesson independently. When a principal or an external observer observed a classroom without the other present, we call these *solo observations*.

- A teacher's final evaluation is the result of the summation of one's individual ratings of various formal observations throughout the year on each of the following four domains of TPAS:

Domain 1: Planning and Preparation

Domain 2: Classroom Environment

Domain 3: Instruction

Domain 4: Professional Responsibilities

- The researcher was also interested in whether or not teachers' scores on specific rubric indicators changed over time. However, the school system does not include a single final rating at the indicator level. Instead, each time a teacher received a formal observation, every indicator received a categorical rating. There are four category choices in the school system called domains on which teachers are rated. In order to construct an overall annual rating on specific instructional indicators, the researcher first converted each categorical rating to an integer, with the lowest possible ratings converted to a 1, the second lowest converted to a 2, and so on. The researcher then averaged each teacher's ratings from the school year in that indicator to obtain a value between 1 and 4. Based on that final average, the researcher assigned the following labels:

- • *"Ineffective"* - Averages less than or equal to a 2.
- • *"Developing"* - Averages greater than a 2 but less than a 3.
- • *"Effective"* - Averages equal to or greater than a 3 but less than a 3.5.
- • *"Highly Effective"* - Averages equal to or greater than a 3.5.

Survey Pilot Test. The researcher asked five individuals to pilot the survey that had acute familiarity with the local teacher evaluation system but were not included in the target population. These respondents were asked to take the survey and agreed, knowing their data would be erased and not included in any part of the study's data collection. Each individual received an email with a link to the survey along with the cover letter. The researcher requested feedback on five areas of the survey:

- Readability
- Grammar/Syntax
- Efficiency
- Organization
- Amount of time spent on the survey

All five participants responded to the survey and edits were made based on the pilot test feedback. From the feedback, the surveyor revamped a number of questions so as to mitigate some ambiguity in the query. The surveyor also changed the order of the questions for purposes of thematically collating said question into more similar themes.

Procedures/Data Collection Methods. Using the University of Maryland's Qualtrics software, the research utilized a quantitative design consisting of an online survey. The Qualtrics survey centered on the district's tenured teacher perceptions of the teacher evaluation system (TPAS) presently utilized across the school system and its impact on the quality of their teaching. As part of this survey, participants also self-reported the frequency with which teachers have engaged in self-reflective behaviors and the level of confidence that such behaviors impacted their instructional delivery. Teachers answered various questions

that utilized a Likert rating scale and self-report demographic information such as if they are a NBC teacher (or not), the number of teaching years, etc. To develop the relevant questions for the survey, this researcher consulted a plethora of well-known resources about teacher perceptions of professional development and self-reflection in general such as the "Teacher Evaluation Profile (Duke & Stiggins, 1986) and the Chicago Public Schools 'My Voice, My School Teacher (MVMS) Survey (Levenstein, 2016).

It also should be noted that throughout this phase of the study, the researcher attempted to quantitatively identify any tangible difference in the perceptions of various teacher cohorts (disaggregated by demographic data) towards one's understanding of what it meant to self-reflect, whether or not each cohort had actively participated in any structured or unstructured self-reflection, the frequency thereof, and, ultimately, declarative degrees of confidence in the impact of self-reflective behaviors specifically, with regard to improving one's own teaching practice.

Research Questions	Survey Questions
<p style="text-align: center;">Demographic Data</p>	<ul style="list-style-type: none"> • Q1. How long have you been employed as an educator in your career? • Q2. Are you presently tenured in the St. Mary's County Public Schools (SMCPS) as a teacher? • Q3. Do you presently hold credentials as a National Board Certified Teacher (NBCT) in your field of expertise? • Q4. Are you currently in the process of pursuing National Board Certification (NBC) in your field of expertise? • Q5. What is the highest degree that you have been conferred? • Q6. Which of the following would best describe you as a teacher? • Q7. How long have you taught at your current school? • Q8. Have you changed schools at all over the last three consecutive school-years? • Q9. Has your local administrative team (principal and/or assistant principal) changed over the last three years? • Q10. Has your content supervisor changed over the last three years? • Q11. Approximately, how many times in the current school year have you been formally observed using the Teacher Performance Assessment System (TPAS) in SMCPS by at least one administrator? • Q12. Approximately, how many times in your entire career in SMCPS have you been formally observed using the Teacher Performance Assessment System (TPAS) by at least one administrator? • Q13. To the best of your knowledge, what was your overall TPAS rating on the very first formal observation you ever received from an administrator in SMCPS? • Q26. Generally speaking, which of the following most accurately reflects your most recent TPAS rating you have received?

Figure 5. Maps the research questions with the survey questions

Research Questions	Survey Questions
<p>1. Do tenured teachers believe that the current teacher evaluation system used in their district – the Teacher Performance Assessment System (TPAS) - impacts the quality of their teaching?</p>	<ul style="list-style-type: none"> • Q17. As an educator, what statement below best describes your professional opinion about the TPAS process and using TPAS as a tool for teacher evaluation? • Q18. In your professional opinion, what has been the impact of the TPAS process on your development as a classroom teacher? • Q19. In your professional opinion, which part of the TPAS process has positively impacted your teaching the most as a teacher in SMCPS? • Q21. In your professional opinion, which of the following descriptors best describe your experience of the TPAS process up to this point as a classroom teacher in SMCPS? • Q22. In your professional opinion, which of the following descriptors best describe the ideal state of the TPAS process as it could be as a classroom teacher in SMCPS? • Q23. How would you rate the TPAS process in your development as a teacher with respect to the following professional duties?
<p>2. Do tenured teachers believe that being evaluated in the Teacher Performance Assessment System (TPAS) and its four component domains improve teachers' effectiveness over time?</p> <p><i>a. Domain I: Planning and Preparation</i></p> <p><i>b. Domain II: Instruction</i></p> <p><i>c. Domain III: Classroom Environment</i></p> <p><i>d. Domain IV: Professional Responsibility</i></p>	<ul style="list-style-type: none"> • Q14. In your professional opinion, which singular TPAS Domain most directly impacts your teaching skill? • Q15. In your professional opinion, which singular TPAS Domain is least likely to impact your teaching skill? • Q16. In your professional opinion, can you rank order below each TPAS Domain that impacts your teaching skill from least effective to most effective?

Figure 5 (cont.). Maps the research questions with the survey questions

Research Questions	Survey Questions
<p>3. To what extent do tenured teachers engage in self-reflection about their teaching?</p>	<ul style="list-style-type: none"> • Q20. In your professional opinion, how much of the TPAS process engages a teacher in behaviors that are self-reflective? • Q27. Did you participate in any of the following self-reflective behaviors, and what was the impact of these self-reflective behaviors on your development as a teacher? • Q28. Did you participate in any of the following self-reflective behaviors, and how often did you personally engage in these self-reflective behaviors? • Q29. As a teacher, I generally know how to "self-reflect" on my own practice.
<p>4. Do tenured teachers that have been evaluated in the Teacher Performance Assessment System (TPAS) grow in teacher effectiveness over time (as measured by TPAS ratings) and, if so, what factors account for this growth?</p>	<ul style="list-style-type: none"> • Assorted demographic data such as Q1(years of teaching); Q3(NBCT or not); Q5(highest degree conferred); Q6(placement level of teacher); Q12 (total number of TPAS observations in career); Q13(first TPAS rating); Q26(last TPAS rating) • Q24. Which of the following professional development activities has positively impacted your instruction the most? • Q25. As a teacher, how do you see yourself spending most of your "learning time" in the schoolhouse?

Figure 5 (cont.). Maps the research questions with the survey questions

Hypothesis. The researcher's hypothesis was encompassed both teacher perceptions of the school system's evaluation tool and the impact of teachers engaging in various self-reflective activities. Positive teacher perception of the Teacher Performance Assessment System (TPAS) was thought to be somewhat muted given that the deployment of TPAS across the county is, on occasion, inconsistent, focused on compliance/quality control, and the system lacks infrastructural capacities to ensure full fidelity to the System. Moreover, various teacher demographic cohorts would have differing degrees of confidence in TPAS given their various experiences with TPAS. Lastly, teachers who willingly engage in more formalized self-reflection and self-assessment yield higher degrees of teacher effectiveness as measured on a TPAS, no matter a teacher's demographic cohort.

Human Subject Review and Confidentiality. All appropriate steps in receiving approval was taken before conducting the research. Participants had no known risks for involvement in this study. A review and analysis of the survey responses and records reflection did not cause any harm or risk to any teacher in the district. Great care was taken to ensure the identities of all participants and were held with the utmost confidentiality. Participation in the survey was completely anonymous and voluntary. Participants did not experience any greater risk than they would typically encounter in their daily work-life. The researcher also made sure his role in the district, for purposes of this study, was understood and was willing, if necessary, to employ a third party to disseminate the surveys. The researcher was clear in explaining the purpose of the study, gaining consent, and sharing the security measures to ensure confidentiality. The researcher also went to great lengths to report honestly and within the appropriate guidelines so as not to disclose evidence that would either harm participants or identify them specifically. To protect the secondary school, the individuals in the study, and

the University of Maryland, the researcher adhered to the guidelines for the University of Maryland Institutional Review Board (IRB). The following procedures were used to ensure that the identities of all respondents remained confidential.

- All participants received a letter describing the study and detailing their confidentiality in participation in the study.
- All participants agreed with an informed consent electronically before beginning the survey.
- To maintain confidentiality, the researcher did not use names or identifying information in the survey results. Survey respondent names were coded into ID numbers and were referred to as an ID number during the analysis.
- The encoded data and all identifying information were removed prior to the researcher analyzing the results.
- Results were reported in aggregate form (by position, gender, advanced, novice, etc.) to protect the identity of the participants.
- The research retained the data from the surveys electronically on a personal computer, and the researcher has sole access to the information contained on said computer.
- Individual data was not shared with any other individuals in the school.
- All data will be erased after 5 years.

Summary. Through this study, the researcher will hope to calibrate teacher perceptions – and confidence - about the local teacher evaluation metric (TPAS), drilled down to each component TPAS Domain. Additionally, this study will investigate tenured teacher cohorts’ perceived degrees of engagement in more formalized practices of both self-reflection and self-assessment and examine its collective impact on teacher quality as compared to other differentiated teacher cohorts in the local district. Notwithstanding, the researcher hopes to quantify if teachers, in general, believe that engaging in self-reflecting behaviors positively impact their own teaching practice. The thought is that teacher cohorts that actively engage in more self-reflective behaviors summarily yield higher ratings for those cohorts on a local district’s teacher evaluation system (TPAS). This is significant since on the district’s teacher evaluation system there is limited explicit embedded/formal self-reflection processes in which teachers can engage. This analysis will help school leaders to make a more concerted effort in how they can integrate processes of self-reflection throughout a teacher’s evaluation system as well as embed such practice as a staple in their daily routine and repertoire.

Section 2 provided an overview of the methodology for this study that was intended to investigate the tenet that as teachers become more apt to “self-reflect” on their own practice, they are more likely to grow as professionals, yielding higher ratings in any teacher evaluation model and thus becoming an accomplished, “highly-effective” practitioner. This researcher hopes to validate the work of Goldstein and Noguera (2006) that those particular school systems that construct a more thoughtful approach to teacher evaluation - one that engages teachers in self-reflection - yields

benefits far beyond the important goal of quality assurance (Goldstein & Noguera, 2006). The participants, setting, and procedures were also discussed. Specifics regarding the survey and the interviews were also described. Finally a brief description for how the data were analyzed was also provided.

SECTION III: RESULTS

Introduction

This chapter details the results of this investigation based upon an analysis of the data gathered from the teacher survey. The chapter begins by presenting the results based upon the crosswalk of the research questions and the survey questions and, subsequently, offering a brief discussion of (a) the data that specifically responds to each of the research questions and (b) the results of the investigation.

The purpose of this study was to analyze *tenured* teacher perceptions of a local teacher evaluation system, the extent to which teachers engage in self-reflection about their teaching, and to determine if teachers grow in effectiveness over time using such evaluation system; so non-tenured teachers were not included in the research analyses. Using a 29-question survey with an opportunity for comments, this study served to answer four essential questions circulating around improving teacher effectiveness. The 29 questions encompassed four main categories:

1. General teacher perceptions of a local teacher evaluation system
2. Teacher perceptions for each of the four component domains of such evaluation system
3. The extent to which teachers engage in self-reflective activities about their teaching
4. The factors and professional development activities that improve teacher effectiveness over time

A total of 608/1420 participants responded to the survey which resulted in a 42.8% response rate to all 29 questions. This study was conducted in July/August of 2017 and teachers were able to complete the survey within a two-week period on a mobile device, tablet, or desktop computer.

Results

In order to answer the four questions, the researcher examined the demographic and respondent data for purposes of identifying themes and correlations - all the while - collating teacher perceptions of a local evaluation system. In many instances, datum was cross-tabulated to investigate how demographic characteristics influenced teacher perceptions. Identifying the correlations and themes benefitted the researcher by understanding areas of strengths in a local teacher evaluation system, areas for improvement, and operative next steps for how best to yield higher levels of teacher effectiveness over time. These various analyses can also aid school administration and/or future researchers in identifying what specific professional development activities most actionably impact teacher quality.

Demographic Data

Demographic data was collected to help identify trends in technology use based on years of teaching, highest degree conferred, National Board Certification, and leveled teaching assignment (elementary, middle, or high school) from only *tenured* teachers. Figure 6 below details the crosswalk of demographic data with survey questions. All demographic data emanated from SMCPS teacher respondents to Survey Questions 1, 2, 3, 4, 5, 6, 12, 13, and 26, respectively. For the purposes of the specificity of this research, Survey Questions 7, 8, 9, 10, and 11 were omitted from the investigation.

Research Questions	Survey Questions
<p style="text-align: center;">Demographic Data</p>	<ul style="list-style-type: none"> • Q1. How long have you been employed as an educator in your career? • Q2. Are you presently tenured in the St. Mary's County Public Schools (SMCPS) as a teacher? • Q3. Do you presently hold credentials as a National Board Certified Teacher (NBCT) in your field of expertise? • Q4. Are you currently in the process of pursuing National Board Certification (NBC) in your field of expertise? • Q5. What is the highest degree that you have been conferred? • Q6. Which of the following assignment would best describe you as a teacher? • Q12. Approximately, how many times in your entire career in SMCPS have you been formally observed using the Teacher Performance Assessment System (TPAS) by at least one administrator? • Q13. To the best of your knowledge, what was your overall TPAS rating on the very first formal observation you ever received from an administrator in SMCPS? • Q26. Generally speaking, which of the following most accurately reflects your most recent TPAS rating you have received?

Figure 6. Crosswalk for Demographic Data and Survey Questions

Years of Teaching Experience

For the school system involved in the study, there is a demonstrable veteran teacher demographic employed with close to 50% of respondents having more than 16 years of teaching experience as a teacher. Specifically, 25.86% of respondents had more than 20 years of teacher experience and over 23% (i.e., 23.69%) of survey respondents had taught between 16 and 20 years. 21.34% of teachers who had completed the survey had between 11 and 15 years of experience coupled with 18.44% of respondents having 6 to 10 years of experience. Respondents with 3 to 5 years of teaching experience comprise 10.67% of the target population. Figure 7 delineates the years of teaching experience for all teacher respondents.

Years of Teaching Experience

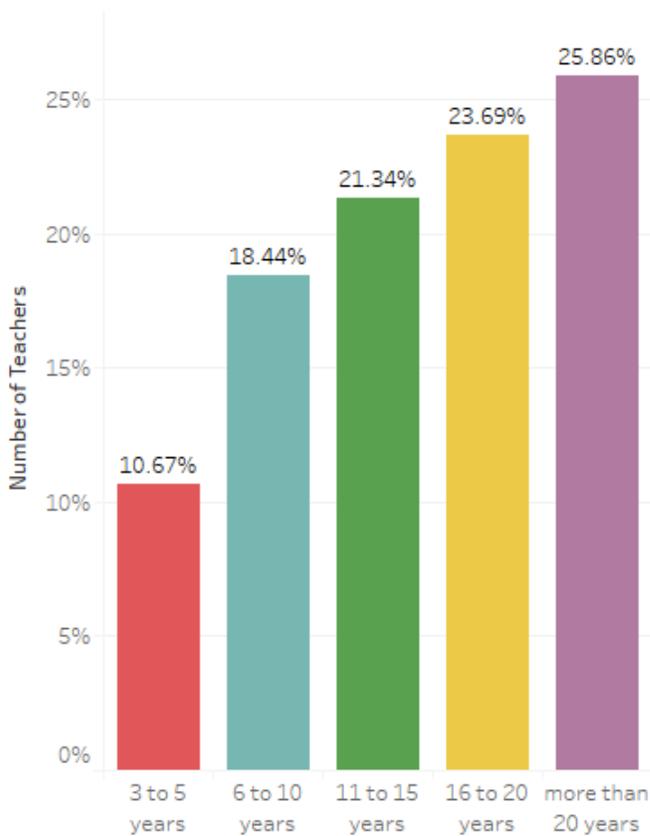


Figure 7. Years of Teaching Experience (by respondent percentage)

Highest Degree Conferred

Figure 8 below conveys the highest degree conferred for all teacher respondents. Teachers holding a Master's degree overwhelmingly comprised close to 80% of total respondents with 74.09% and those specifically holding Master's degrees in education. Teachers with doctorates populated 1.38% of total respondents while 13.64% of teachers were conferred a Bachelor's degree in education vice those 5.18% of teachers holding a Bachelor's in their area of expertise (e.g., B.S., in Mathematics).

Highest Degree Conferred

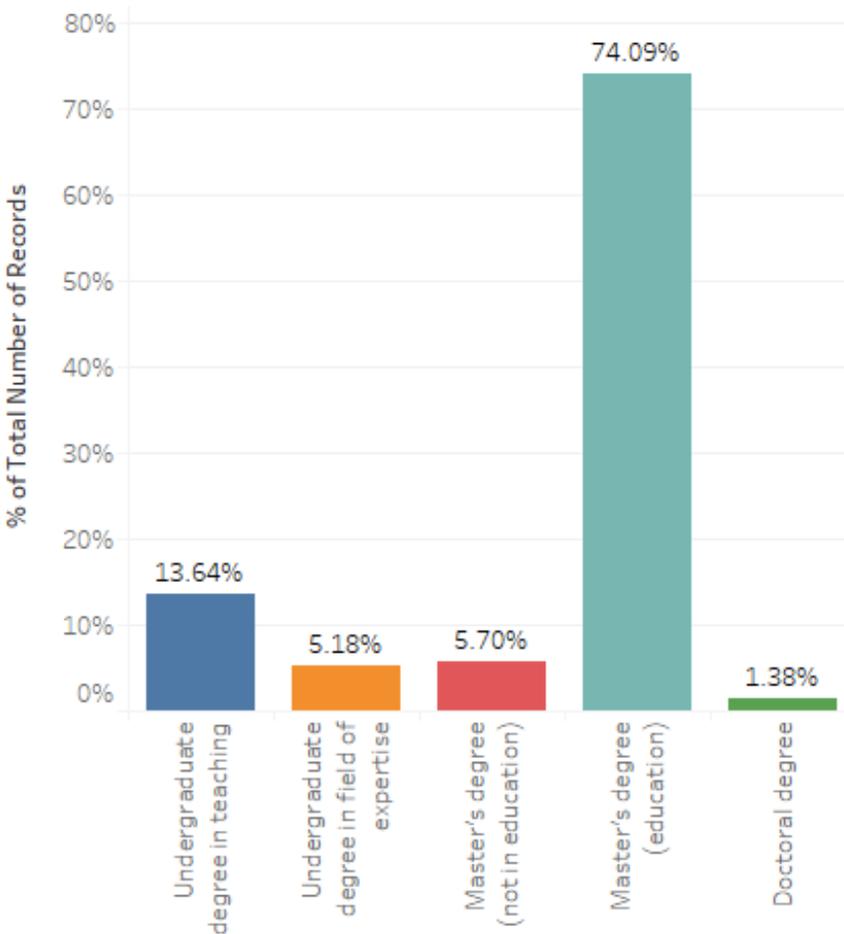


Figure 8. Highest Degree Conferred (by respondent percentage)

NBCT vs. non-NBCT

One of the focal points for this study was to identify the cohort of Nationally Board Certified teachers (NBCT) and juxtapose their first/last self-reported TPAS rating(s), perception of the TPAS system as impactful (or not) on improving teacher effectiveness, and, their propensity to engage in self-reflective activities against those tenured teachers from the local school system that were not National Board Certified (non-NBCT). From the survey respondents, there were 36 NBCT from the school system, which comprised 6.22% of the total respondents; hence, close to 94% of teachers are non-NBCT; Figure 9 below encapsulates the percentage of each teacher cohort in the local school system.

NBCT vs. non-NBCT Cohort

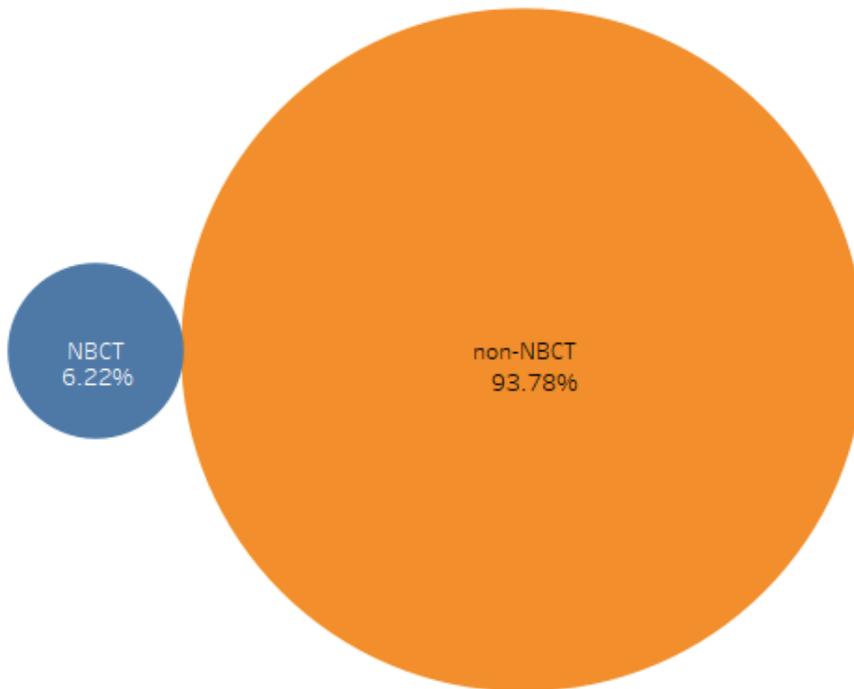


Figure 9. NBCT vs. non-NBCT Cohort (by respondent percentage)

Leveled Teaching Assignment

Rounding out the demographic data for the survey were the leveled teaching assignments, with over 50% of all respondents self-reporting that they were assigned to an elementary school (54.96%). As seen in Figure 10, approximately 16% of teachers reported to have worked in a middle school (specifically, 15.98%) and 29.06% of teachers responding to the survey were high school teachers.

Teaching Assignment - by level

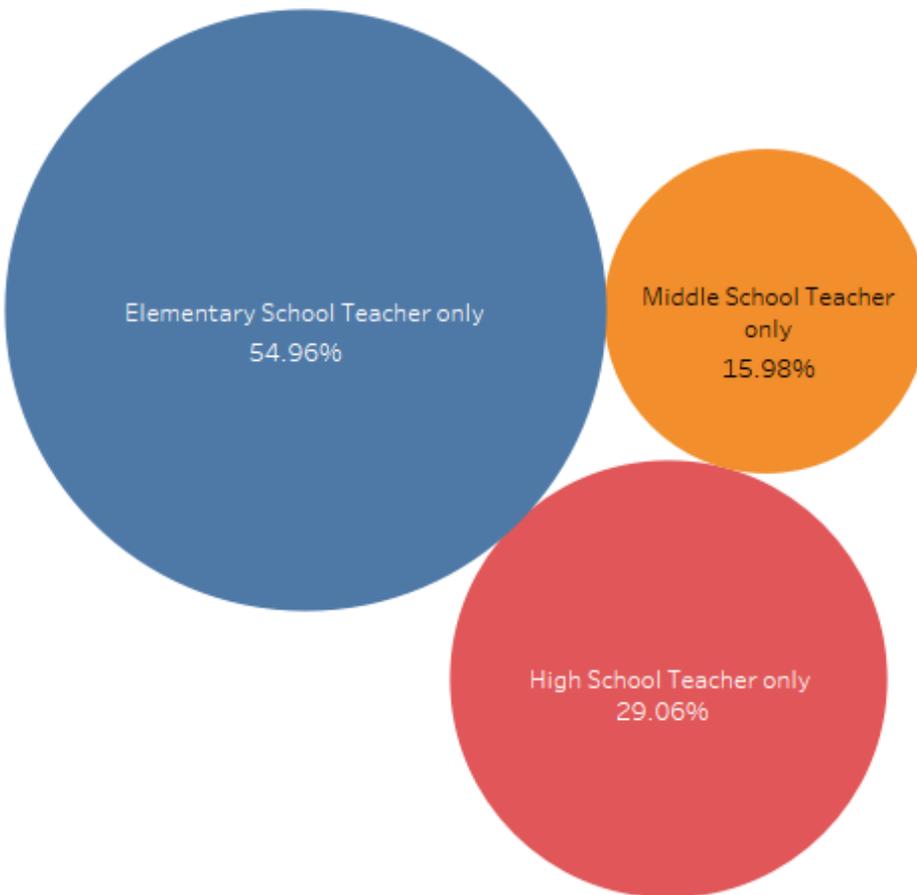


Figure 10. Leveled Teaching Assignment (by respondent percentage)

Total Number of Career TPAS Observations

As previously stated, close to 50% of all teachers reported to have been teaching for more than 16 years; so there would have been a lot of opportunities for teachers in the local school system to have been formally evaluated within TPAS. Hence, the survey request to query respondents to approximate the total number of formal TPAS observations that they have experienced in their career. This request is grounded in the fact that teacher evaluation systems like TPAS profess to be formative. That is, the more a teacher is evaluated within a particular evaluation system, the prevailing thought would be that a teacher would improve their performance over time using that evaluation system the more times that they are formally observed. In essence, there should be a direct, positive relationship between the number of TPAS observations experienced and one's TPAS rating. Figure 11 encapsulates the total number of TPAS observations one has experienced over a career that were reported via the survey. Approximately 25% of teachers participating in the survey reported to have been observed more than 10 times over their tenure. Close to 45% of all respondents (44.90%) have been formally observed using TPAS between 6 and 10 times; the next most popular response was 23.14% of teachers self-reporting to have been observed between 3 and 5 times over their career. Only 6.67% of all teachers responded that they have been observed 1 to 2 times. As a gentle reminder, only tenured teachers were considered part of the target population of the investigation.

Total Number of Career TPAS Observations

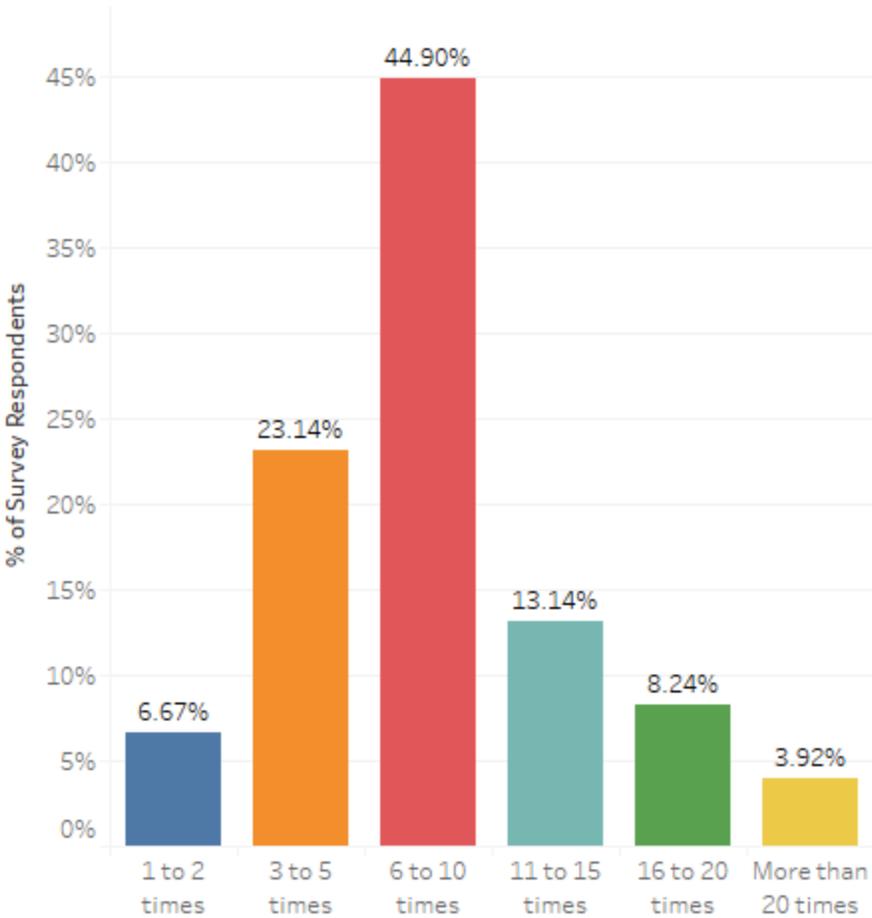


Figure 11. Total number of career TPAS observations (by respondent percentage)

Initial and Most Recent TPAS Rating

Because the study is investigating teacher effectiveness over time, the researcher thought it prudent to get a baseline TPAS rating for respondents in some capacity; more specifically, a teacher’s first and last (i.e., most recent) formal TPAS rating would not only be somewhat memorable but also would give rise to any substantive growth a teacher would experience over their career. Figures 12 and 13 below capture the self-reporting TPAS rating of a teacher’s first and last formal observation.

Self-Reported First TPAS Rating - by percentage

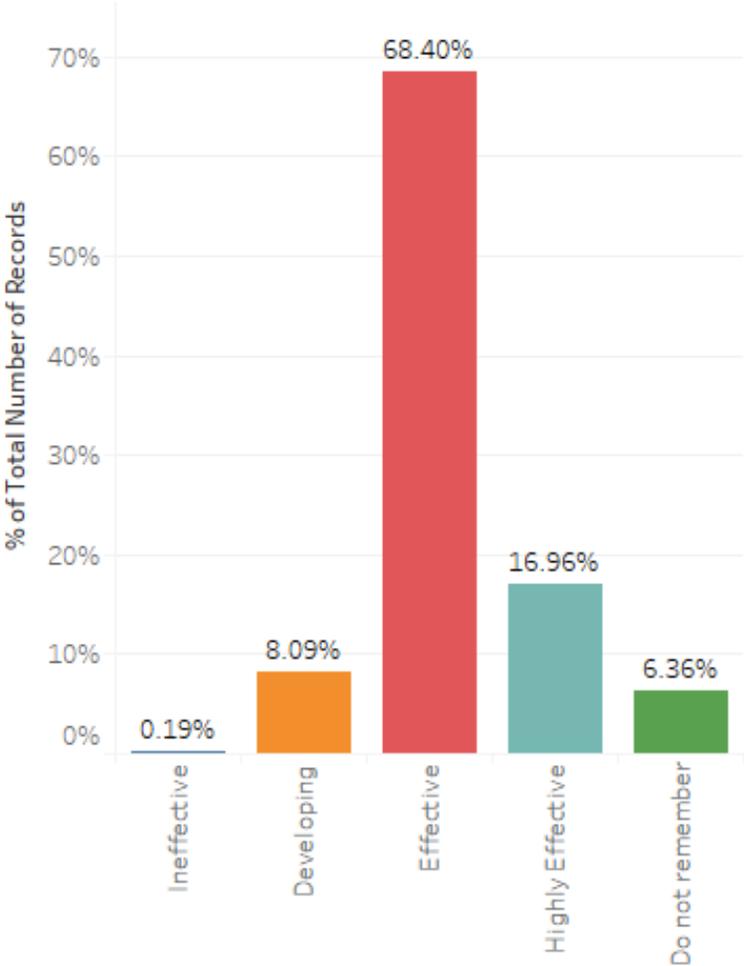


Figure 12. Self-Reported Initial TPAS rating (by respondent percentage)

What is interesting is that close to 70% of all respondents claimed to have been deemed “effective” on their first, formal TPAS rating; which gives credence to many recent criticisms of teacher evaluation systems, such as *The Widget Effect*, that most teachers are evaluated as “effective” no matter their true performance. In essence, schools systems fail to recognize differences in teacher effectiveness and, as a result, refrain from acting on these differences in teacher effectiveness; thus allowing teachers to continue in their stead without the requisite support, guidance, and/or intervention to improve their practice.

Self-Reported Most Recent TPAS Rating - by percentage

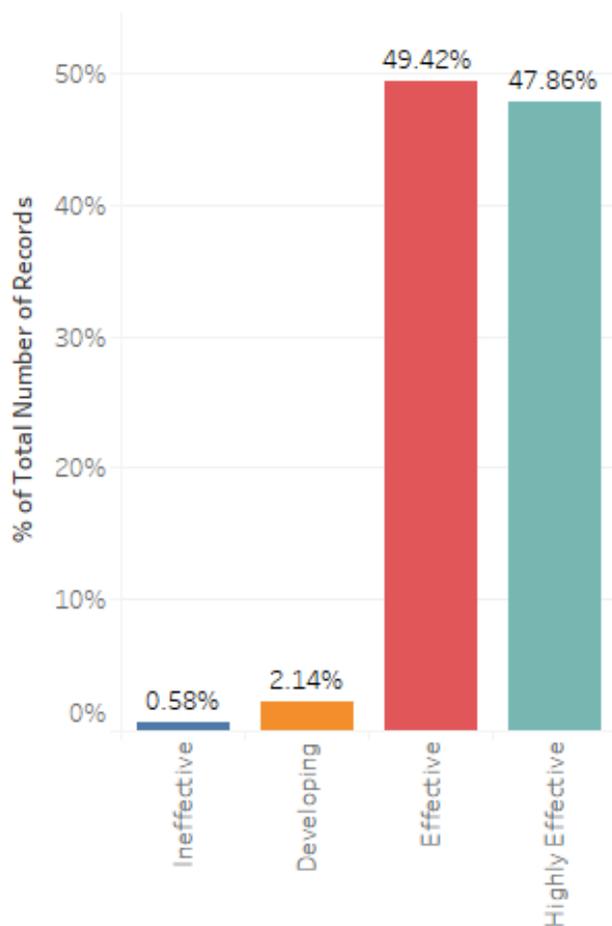


Figure 13. Most Recent Self-Reported TPAS rating (by respondent percentage)

The survey data was reviewed to glean correlations between teacher effectiveness – as reported on TPAS ratings – and various demographic data points. Factors such as years of teaching, National Board Certification, leveled teaching assignment, the total number of TPAS observations experienced, first/last TPAS rating, and/or highest degree conferred, could affect the likelihood of a teacher’s perception of the local teacher evaluation system, propensity to self-reflect on their own classroom practice, and other professional development activities all geared towards improving teacher effectiveness over time. As we proceed through this investigation, the researcher will use multiple variables (in the form of different

demographical cohorts) – toggling back and forth among all of them as appropriate - in the hopes of uncovering viable interrelations and analytical dependencies via charts and cross tabulation tables.

Research Question One: Teacher Perceptions of TPAS

The data related to Research Question One came from SMCPS teacher respondents to Survey Questions 17, 18, 19, 21, 22, and 23 respectively (see Figure 14 below).

Research Question	Survey Questions
<p>1. Do tenured teachers believe that the current teacher evaluation system used in their district – the Teacher Performance Assessment System (TPAS) - impacts the quality of their teaching?</p>	<ul style="list-style-type: none"> • Q17. As an educator, what statement below best describes your professional opinion about the TPAS process and using TPAS as a tool for teacher evaluation? • Q18. In your professional opinion, what has been the impact of the TPAS process on your development as a classroom teacher? • Q19. In your professional opinion, which part of the TPAS process has positively impacted your teaching the most as a teacher in SMCPS? • Q21. In your professional opinion, which of the following descriptors best describe your experience of the TPAS process up to this point as a classroom teacher in SMCPS? • Q22. In your professional opinion, which of the following descriptors best describe the ideal state of the TPAS process as it could be as a classroom teacher in SMCPS? • Q23. How would you rate the TPAS process in your development as a teacher with respect to the following professional duties?

Figure 14. Crosswalk for Research Question 1 and Survey Questions

Let’s begin the TPAS story here by unpacking the respondent data on Q23 and how different demographic cohorts rated TPAS in their development as a teacher with respect to building their own overall teaching capacity. All told, almost 50% of teachers (46.10%)

responded *neutrally* regarding TPAS' impact on building their overall capacity; in fact, every one of the seven different demographic cohorts rated TPAS *Neutral* as its most popular option for Q23 below in Table 1. Most teacher cohorts did in fact rate TPAS somewhere in the middle of the continuum, with the majority of all respondents claiming that they were either *somewhat dissatisfied, neutral, or somewhat satisfied* with TPAS and its ability to help build a teacher's overall capacity. Interestingly enough, the strongest reaction *against* the TPAS process and responded that they were *very dissatisfied* was evidenced in two particular demographics: teachers that had experienced more than 20 rated observations and high school teachers. It should be noted that among that same demographic of teachers with more than 20 rated TPAS observations, close to 9% of these same teachers (8.99%) were very satisfied with TPAS and its ability to build overall capacity. Hence, we can conclude that TPAS elicits the strongest sentiment – both *very satisfied* and *very dissatisfied* – with those teachers who have been formally observed the most with TPAS.

Table 1

Teacher’s Resposnes to Survey Question 23 (in percent)

		Q23: How would you rate the TPAS process in your development as a teacher with respect to building overall teacher capacity?				
		Very dissatisfied	Somewhat dissatisfied	Neutral	Somewhat satisfied.	Very satisfied
Different Demographic Cohorts	NBCT (n=36)	6.06%	12.12%	45.45%	30.36%	6.06%
	Teaching more than 20 years (n=124)	8.08%	9.60%	44.80%	32.00%	4.08%
	Teacher had more than 20 rated TPAS observations (n=61)	10.80%	20.20%	30.37%	29.64%	8.99%
	Non-NBCT (n=547)	6.54%	11.39%	45.99%	30.38%	5.70%
	Elementary School Teacher (n=193)	5.15%	8.25%	45.36%	34.59%	6.70%
	Middle School Teacher (n=63)	3.23%	11.29%	48.39%	30.65%	6.45%
	High School Teacher (n=110)	11.11%	14.81%	50.93%	19.44%	3.70%
	Teachers rated on TPAS as “highly effective” (n=246)	5%	13.33%	39.58%	35.42%	6.67%
	Overall	6.05%	11.35%	46.10%	30.33%	5.68%

By omitting all *neutral* responses from all disaggregated teacher cohorts and by combining responses that were either *very dissatisfied* or *somewhat dissatisfied* together, coupled with combining those responses that were either *very satisfied* or *somewhat satisfied* in the same vein, the overall TPAS perception of all teacher cohorts seems to be *very or somewhat satisfied* with TPAS' ability to build overall teacher capacity as evidenced in Table 2 below.

Table 2

Compartmentalizing Teachers' Responses to Q23 (in percent)

		Q23: How would you rate the TPAS process in your development as a teacher with respect to building overall teacher capacity?	
		Very or somewhat <u>dissatisfied</u>	Very or somewhat <u>satisfied</u>
Different Demographic Cohorts	NBCT (n=36)	30.35%	69.65%
	Teaching more than 20 years (n=124)	33.33%	66.67%
	Teacher had more than 20 rated TPAS observations (n=61)	42.87%	57.13%
	Elementary School Teacher (n=193)	33.2%	66.80%
	Middle School Teacher (n=63)	25.43%	74.57%
	High School Teacher (n=110)	28.12%	71.88%
	Teachers rated on TPAS as "highly effective" (n=246)	52.83%	47.17%
	Overall	33.09%	66.91%

Q17: TPAS as a Tool for Teacher Evaluation

As previously stated by Marzano, teacher evaluation has a two-fold objective: to ensure *quality assurance* and *teacher development*. Q17 gets to the heart of this statement and queries teachers as to their perception of TPAS as a tool for teacher evaluation through this two-fold lens. Figure 15 below encapsulates teacher responses for Q17 and clearly embraces the notion that TPAS is much more than just a mechanism for teacher quality control with only 2.83% of all respondents believing that is the case. Close to 70% of all survey respondents (i.e., 69.84%) perceive TPAS to be more than a *quality control* tool for teachers and that TPAS encompasses a *compliance* factor as well. Of the aforementioned 70% of respondents, 42.11% of all teachers believe that TPAS is not only an instrument for compliance and teacher quality control but also one that assists in the professional development for teachers as well. On the flip side, only about 6% of teacher respondents felt that TPAS was helpful in improving their classroom instructional delivery only

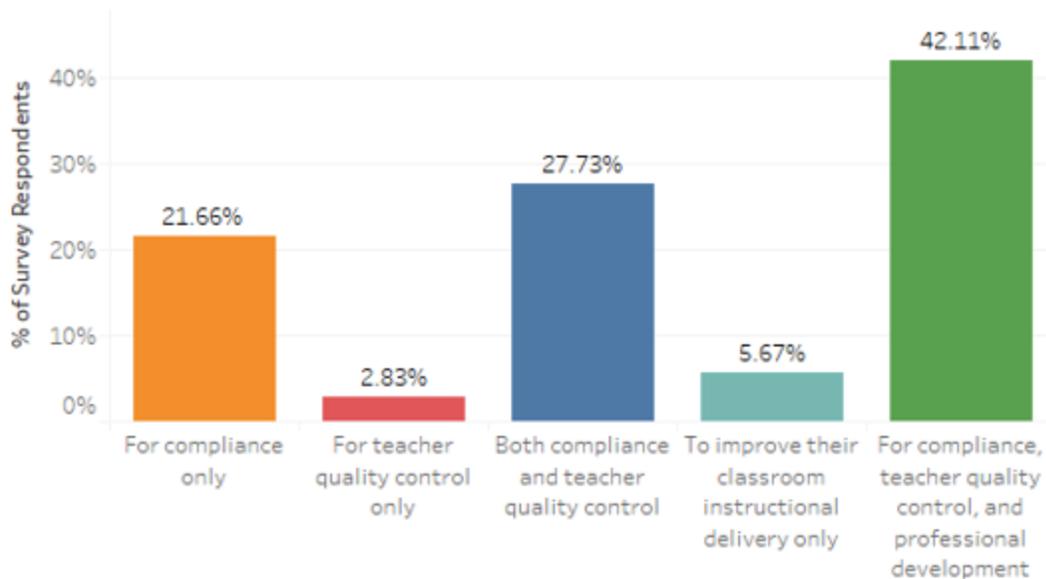


Figure 15. Teacher perception of TPAS as a tool for teacher evaluation

Q18: The Overall Impact of TPAS

Research Question One’s investigative analysis of the survey results moves to Q18 and reviews what teachers feel has been the overall impact of TPAS on own their development as a classroom teacher. Over 20% of all teachers (i.e., 21.27%) self-reported that TPAS had no impact on their development as a classroom teachers and an additional 43.54% of teachers felt that TPAS had only a small impact on their performance. All told, close to 65% of all respondents (i.e., 64.81%) characterized the impact of TPAS on their own teacher effectiveness as either none and/or minimal. Conversely, @35% of teacher respondents felt that TPAS had a moderate to large impact on their development as a classroom teacher; see Figure 16 below for more details.

Q18: What has been the impact of TPAS on development as a classroom teacher?

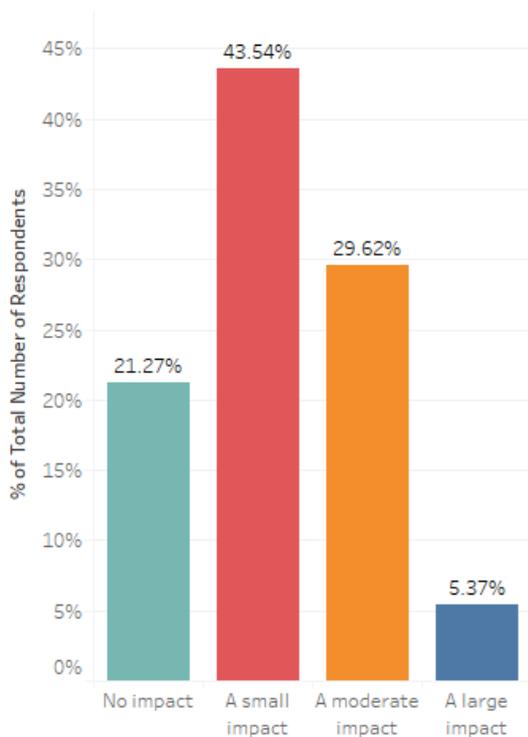


Figure 16. Results from Survey Question 18

In Table 3, the researcher wanted to drill down to each demographic teacher cohort and perform a cross-tabular analysis of Q18 to see if there were any identifiable patterns that would emerge. Was there a particular teacher demographic that presented as an outlier given the Q18? As expected, the majority of each disaggregated teacher cohort responded that TPAS had *no impact* to a *small impact* in their development as a classroom teacher. In fact, over 70% for each of the following five groups' respondents reported that TPAS had zero to minimal impact on their teacher development: NBCT; teachers for more than 20 years; elementary teachers; middle school teacher; and high school teachers. Additionally, *not* one NBCT (i.e., 0.00%) reported that TPAS had a large impact on their teaching acumen. Conversely, the one outlier cohort that reported the largest impact of TPAS (i.e., 16.67%) on their teaching were those teachers that had experienced more than 20 observations in their career; this response was more than 10 percentage points above the overall survey response on Q18 for TPAS having a *large impact* on teaching development (i.e., 5.37%).

Table 3

Teachers' Responses to Survey Question 18 (in percent)

		Q18: In your professional opinion, what has been the impact of the TPAS process on your development as a classroom teacher?			
		No impact	A small impact	A moderate impact	A large impact
Different Demographic Cohorts	NBCT (n=36)	38.89%	47.22%	13.89%	0.00%
	Teaching more than 20 years (n=124)	33.33%	40.65%	23.58%	2.44%
	Teacher had more than 20 rated TPAS observations (n=61)	11.11%	50.00%	22.22%	16.67%
	Elementary School Teacher (n=193)	16.67%	42.71%	35.42%	5.21%
	Middle School Teacher (n=63)	21.31%	49.18%	24.59%	4.92%
	High School Teacher (n=110)	28.70%	42.59%	25.00%	3.70%
	Teachers rated on TPAS as "highly effective" (n=246)	19.41%	45.57%	29.96%	5.06%
	Overall	21.27%	43.54%	29.62%	5.37%

Q19: The Most Impactful Part of the TPAS Process

As we continue to unpack survey data for Research Question 1 with regards to teacher perception of TPAS, the investigation now segues into which component part of the TPAS process has impacted teaching the most? Is it at the “*Pre-Conference*” with the evaluator and evaluatee conferring about the upcoming observation and reviewing the lesson plan? Or is it during the “*lesson-planning*” in which the evaluatee decides how to differentiate the instruction, what embedded educational technology to include in the upcoming lesson, and/or what curricular resources should be utilized for the lesson? In Figure 16 below, teachers overwhelmingly responded (i.e., 67.29%) that the *post-conference* was the most impactful part of the TPAS process on one’s teaching prowess; it was not the *actual observation* (9.35%) nor the *lesson-planning* (17.76%), respectively.

Q19: Which part of the TPAS process has impacted your teaching the most?

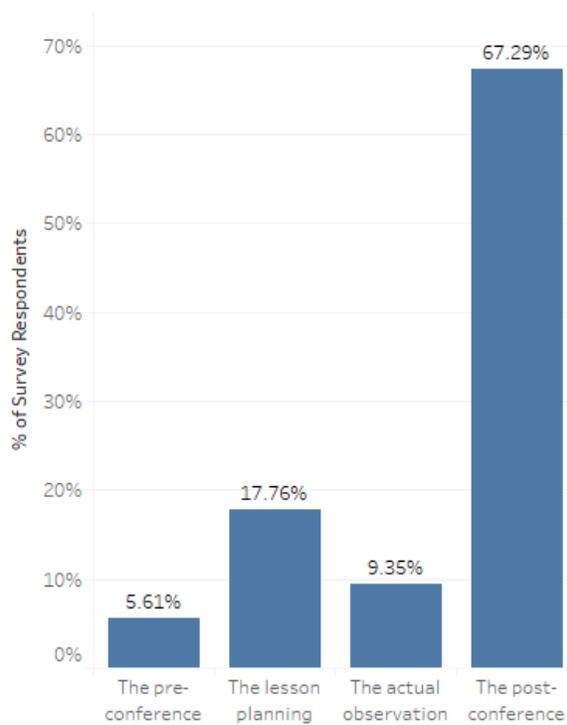


Figure 16. Results from Survey Question 19

As we review the cross-tabulation of Q19's data in Table 4 below, all demographic cohort have responded similarly in kind (with the overall survey respondents) that the *post-conference* was the most impactful part of the TPAS process. Over 60% of all teachers rated as “highly effective” in TPAS (i.e., 64%), elementary teachers (62%), and high school teachers (61%) all responded to the survey that the TPAS *post-conference* had the most positive impact on one's teaching out of all of the assorted components of the TPAS process. All teacher cohorts (except for “highly effective” teachers) responded that the “lesson planning” was the next most impactful part of the TPAS process; “highly effective” teachers responded that the “actual observation” was the 2nd most impactful component of TPAS.

Table 4

Teachers' Responses to Survey Question 19 (in percent)

		Q19: Which part of the TPAS process has most positively impacted your teaching?				
		The Pre-Conference	The Lesson Planning	The Actual Observation	The Post Conference	Other
Different Demographic Cohorts	NBCT (n=36)	0%	27%	13%	58%	2%
	Teaching more than 20 years (n=124)	5%	15%	8%	59%	13%
	Teacher had more than 20 rated TPAS observations (n=61)	4.8%	15.5%	8.1%	58.5%	13.0%
	Elementary School Teacher (n=193)	10%	15%	5%	62%	8%
	Middle School Teacher (n=63)	13%	21%	5%	54%	8%
	High School Teacher (n=110)	5%	17%	9%	61%	8%
	Teachers rated on TPAS as "highly effective" (n=246)	9%	12%	23%	64%	2%
	Overall	8%	18%	6%	60%	8%

Q23: Perception of TPAS w respect to Professional Duties

Teachers were also asked to rate the TPAS process in their development as a teacher with respect to the following professional teaching duties:

- Dealing with the day to day teacher “workflow”;
- Implementing the Common Core;
- Using student data;
- Engaging in self-reflection and self-assessment;
- Integrating educational technology;
- Differentiating instruction;
- Lesson planning and using curricular resources;
- Working in a professional learning community (PLC); and
- Classroom management

Survey respondents were asked to respond to each of the aforementioned teaching duties and record the degree to which each respondent was “satisfied” or “dissatisfied” (including “very satisfied” or “very dissatisfied”) with TPAS and its ability to build the requisite capacity in teachers. Respondents were most positive with TPAS’ ability to “engage teachers in self-reflection”; “lesson planning and using curricular resources”; “differentiating instruction”; and “classroom management.” Respondents also acknowledged that they were “satisfied” with TPAS and its ability to help teachers “implement the Common Core”; “use student data; “work in a PLC”; and “integrate educational technology.” Teachers did not rate TPAS as being particularly effective in helping teachers deal with the “day to day” work flow of the job. See Figure 18 below for more details.

Overall SMCPS Teacher Perception of TPAS and Positive Impact on Professional Duties



Figure 18. Results from Survey Question 23

Q23: How would you rate the TPAS process in your development as a teacher with respect to the following professional duties?

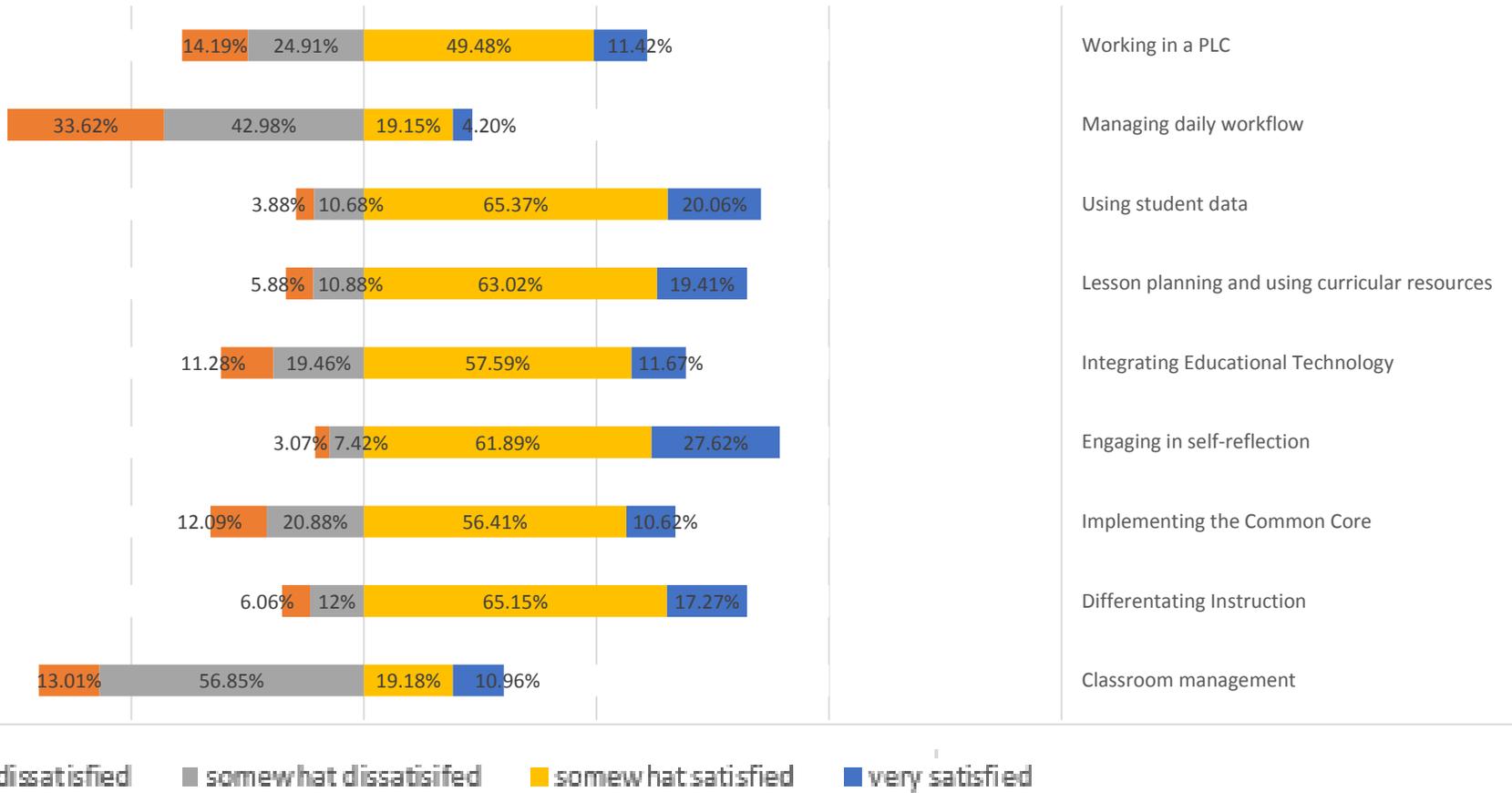


Figure 19. Additional Results from Survey Question 23

Q21 and Q22: Past Perceptions and Future Possibilities of TPAS

As we finish up Research Question 1's analysis of teachers' perception of TPAS on an array of topics, the researcher thought it would be interesting to juxtapose past teacher experiences with TPAS and what teachers thought TPAS "could be" in an ideal sense (if deployed to fidelity). As Danielson had previously stated about her own Framework (of which TPAS is a derivative), her intent was always to implement TPAS *formatively* - not *summatively* – so as to assist teachers in improving their various aspects of their teaching acumen. In the consecutive figures (i.e., Figures 20 and 21) below, the researcher has first collated responses of teacher perception of their past experiences with TPAS (Figure 21) and, following, identified what TPAS possibilities could exist in the future if TPAS is able to be tweaked a bit in its utilization (Figure 22).

Surprisingly, 46.03% of all teachers most often reported that past TPAS experiences facilitated a "*reflection*" of some capacity with regards to their teaching. The second most popular response from teachers was that TPAS – in their opinion - was a "poor use of time"; over 20% of all respondents (21.66%) responded with that perception. Only 1.55% of all teachers responded that their past experiences with TPAS were energizing. Other rather low respondent marks for past TPAS experiences included TPAS being characterized as impactful (2.71%); and/or having a lack of *engagement* (5.03%).

Past Teacher Experience with TPAS

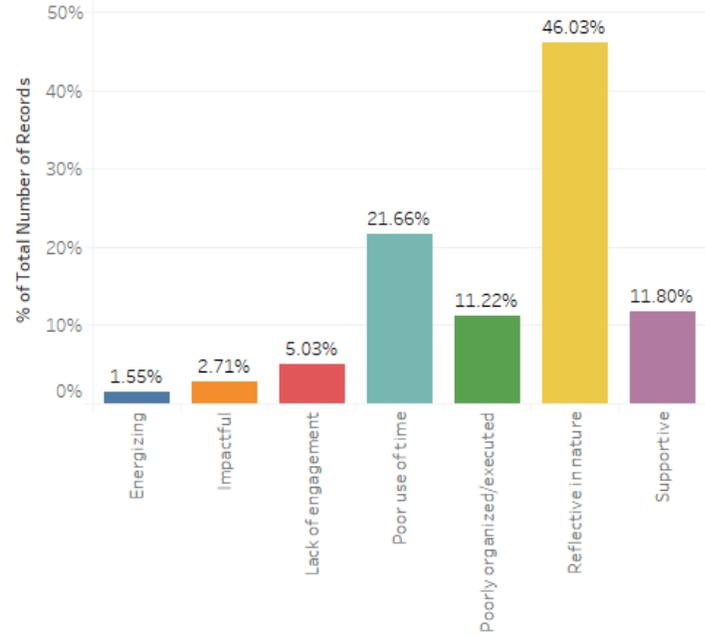


Figure 20. Past Teacher Experience with TPAS

TPAS Possibilities in the Future: What TPAS "could be"

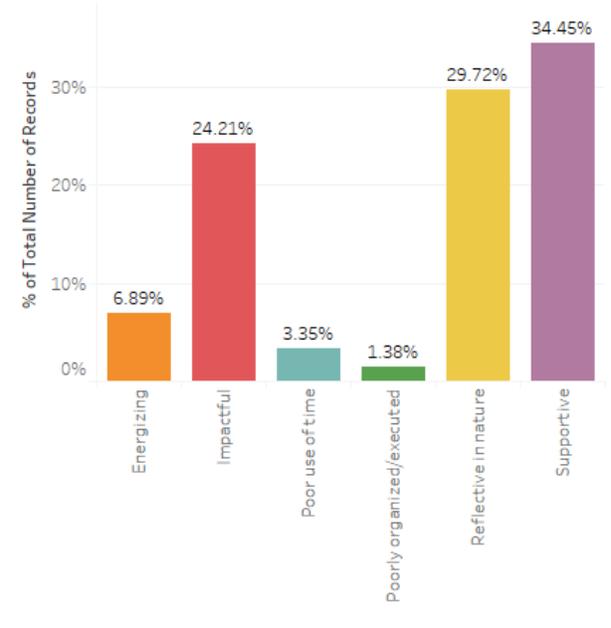


Figure 21. TPAS possibilities in the future

As the survey continued, teachers were then asked in Q22 about what TPAS *could actually look like* if implemented with fidelity across the school system. The Q21 data juxtaposed with Q22 somewhat validates this implementation divide between what Danielson had intended for her (evaluation) Framework and the reality of what is actually happening in school houses. As we review Table_ and the delta values between past and future/”what if” characterizations of TPAS the researcher perseverated on the greatest deltas between the two columns; this would represent the largest difference in teacher perception between TPAS *reality* and TPAS as *it could be* in its ideal state.

The greatest divide of TPAS characterizations resided with the notion that TPAS was “*supportive*”, encompassing an absolute delta value of 22.5% from all respondents; the next largest delta value for teachers was TPAS being characterized as “*impactful*” ($\Delta = 21.6\%$) between *past* experiences ($\Delta=2.7\%$) of TPAS with what TPAS *could be* in the future ($\Delta=24.3\%$). “*Poor use of time*” ($\Delta = 18.3\%$), and TPAS being “*reflective in nature*” ($\Delta = 16.3\%$) represented the next two greatest divides for teacher respondents. The smallest delta value for survey respondents resided with the characterization that TPAS “*lacked engagement*” ($\Delta = 5.1\%$), implying that present TPAS implementation is closest to this reality and that the local evaluation system actually is “*engaging*” for teachers.

Table 5

Characterizations of Past and Future TPAS Experiences (in percent)

TPAS Characterizations:	<u>Past</u> TPAS Experiences	What TPAS <u>could be</u>	Delta ▲	Greatest ▲
<i>Lack of engagement</i>	5.1%	0.0%	5.1%	7 th
<i>Poor use of time</i>	21.7%	3.4%	18.3%	3 rd
<i>Poorly organized/executed</i>	11.2%	1.4%	9.8%	5 th
<i>Reflective in nature</i>	46.1%	29.8%	16.3%	4 th
<i>Energizing</i>	1.5%	6.9%	5.4%	6 th
<i>Supportive</i>	11.8%	34.3%	22.5%	1 st
<i>Impactful</i>	2.7%	24.3%	21.6%	2 nd

Research Question Two: Teacher Perception of TPAS Component Domains

The data related to Research Question Two came from SMCPS teacher respondents to Survey Questions 14, 15, and 16, respectively.

Research Question	Survey Questions
<p>2. Do tenured teachers believe that being evaluated in the Teacher Performance Assessment System (TPAS) and its four component domains improve teachers’ effectiveness over time?</p> <ul style="list-style-type: none"> • <i>Domain I: Planning and Preparation</i> • <i>Domain II: Instruction</i> • <i>Domain III: Classroom Environment</i> • <i>Domain IV: Professional Responsibility</i> 	<ul style="list-style-type: none"> • Q14. In your professional opinion, which singular TPAS Domain most directly impacts your teaching skill? • Q15. In your professional opinion, which singular TPAS Domain is least likely to impact your teaching skill? • Q16. In your professional opinion, can you rank order below each TPAS Domain that impacts your teaching skill from least effective

Figure 22. Crosswalk for Research Question 2 and Survey Questions

Q14, Q15, and Q16: Most and Least Impactful TPAS Domains

In examining the figures (i.e., Figures 23 and 24) below, the researcher wanted to drill down to tenured teacher perception about the four component domains of the local teacher evaluation system (TPAS). Specifically, the researcher wanted to query teacher respondents regarding which of the four aforementioned TPAS domains were most *and* least impactful in

their professional opinion; especially when it came to affecting their own personal teacher effectiveness. Subsequently, the survey also asked respondents to rank order each of the domains from least to most effective once again with respect to impacting teacher quality.

For these survey questions pertaining to Research Question 2, there were over 500 respondents ($n=501$). Overwhelmingly, 84.11% of teachers responded that TPAS' 4th Domain – *Professional Responsibility* – was the least impactful Domain with regards to substantively affecting their own teaching practice; this response was consistent across every demographic in this study. This certainly seems plausible given that most of the sub-components in Domain IV deal with *Communicating with families* (Component 4b) and *Maintaining accurate records* (Component 4e) since these teacher attributes are not reflected in their explicit instruction.

Teacher Perception: Least Impactful TPAS Domain ($N=501$)

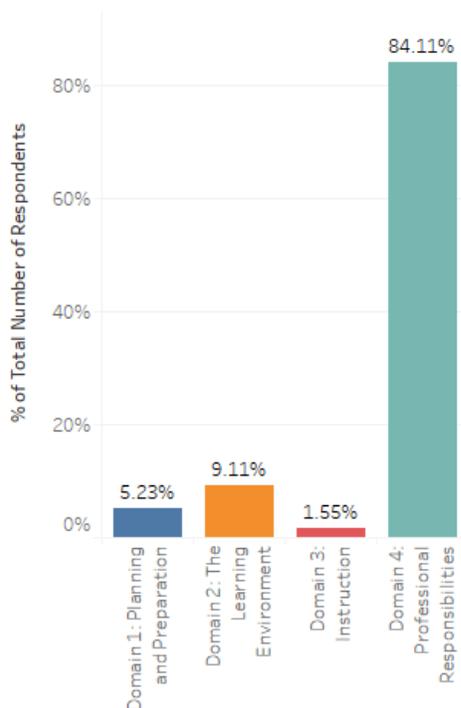


Figure 23. Teacher perceptions of least impactful TPAS Domain

In Figure 24 below, over 50% of all teachers (i.e., 54.46%) self-reported that Domain 3 – Instruction – was the most impactful TPAS Domain with regards to improving their own teacher effectiveness. In reviewing the cross-tabulated demographic data, some teacher cohorts reported higher levels of confidence in Domain 3 and its impact on teacher quality. That is, close to 60% of elementary school teachers, teachers that have taught for more than 20 years, and the NBCT’s cohort all responded with higher degrees of confidence in TPAS Domain 3 than their peers. Only 43% of all middle school teachers reported that the *Instruction* Domain was most important in improving their teaching skill.

Teacher Perception: Most Impactful TPAS Domain (*N* = 501)

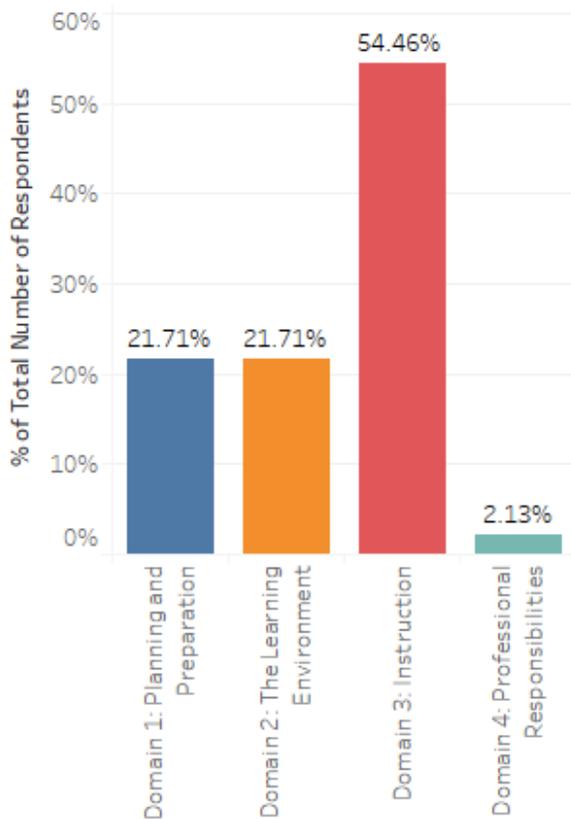


Figure 24. Teacher perceptions of most impactful TPAS Domain

Interestingly enough, 21.71% of respondents – overall - equally felt that the *Planning and Preparation* (Domain 1) and *The Learning Environment* (Domain 2), respectively, were the next most impactful TPAS Domains. Notwithstanding, the researcher noted that there were acute differences in some of our disaggregated cohorts responses to what they reported as the 2nd most impactful TPAS Domain. Teachers rated as “*highly effective*”, elementary school teachers and teachers who have taught more than 20 years responded that the *Planning and Preparation* Domain was the next most impactful TPAS Domain. NBCT was the only demographic cohort that rank ordered Domain 2 - *The Learning Environment* – as the next most impactful Domain after *Instruction*.

Table 6

Teachers' Responses to Survey Questions 14 (in percent)

		Q14: In your professional opinion, which singular TPAS Domain most directly impacts your teaching skill?			
		Domain 1: Planning and Preparation	Domain 2: The Learning Environment	Domain 3: Instruction	Domain 4: Professional Responsibilities
Different Demographic Cohorts	NBCT (n=36)	15%	26%	59%	0%
	Teaching more than 20 years (n=124)	22%	18%	57%	3%
	Elementary School Teacher (n=193)	20%	19%	60%	1%
	Middle School Teacher (n=63)	21%	22%	43%	14%
	High School Teacher (n=110)	24%	24%	50%	2%
	Teachers rated on TPAS as “highly effective” (n=246)	22%	18%	57%	3%
	Overall	21.71%	21.71%	54.46%	2.13%

Research Question Three: Are Teachers Self-Reflective Practitioners?

The data related to Research Question Three came from SMCPS teacher respondents to Survey Questions 20, 27, 28, and 29 respectively.

Research Question	Survey Questions
<p>3. To what extent do tenured teachers engage in self-reflection about their teaching?</p>	<ul style="list-style-type: none"> • Q20. In your professional opinion, how much of the TPAS process engages a teacher in behaviors that are self-reflective? • Q27. Did you participate in any of the following self-reflective behaviors, and what was the impact of these self-reflective behaviors on your development as a teacher? • Q28. Did you participate in any of the following self-reflective behaviors, and how often did you personally engage in these self-reflective behaviors? • Q29. As a teacher, I generally know how to "self-reflect" on my own practice.

Figure 25: Crosswalk for Research Question 3 and Survey Questions

Q20: TPAS Process and Self-Reflection

Research Question Three focused on the extent to which teacher cohorts engage in self-reflection. Questions such as *do teachers participate in self-reflection, the degree/frequency with which cohorts most actively engage in self-reflection, and the confidence that teacher cohorts have in self-reflection as it pertains to impacting their instructional performance* have been addressed. The study reviewed the following teacher cohorts throughout this investigation in order to glean any substantive patterns in a teacher’s propensity to “self-reflect”:

- Leveled teachers (such as elementary, middle, and high school teachers)
- NBCT
- Teachers with more than 20 years of experience
- Teachers that have experienced more than 20 formal TPAS observations
- Teachers that are rated as “highly effective” on their most recent, formal TPAS observation
- Occasional other teacher cohorts were integrated into the study

In Figure 26 below, Q20 captures the teachers’ overall percentage response to the survey question regarding the extent to which the TPAS process - in general – engages a teacher in self-reflection; and the results, quite frankly, were surprising in a good way for this researcher. That is, over 50% of all respondents (54.74%) – agnostic of teacher cohort – reported that the TPAS process engages a teacher in either a “moderate amount” or a “great deal” of self-reflection. 38.88% of teachers felt that the TPAS process engages a teacher in self-reflective activities “a little” while only 6.38% of respondents felt that TPAS does not engage teachers in any amount of self-reflection.

Q20: How much of the TPAS Process engages a teacher in self-reflection?

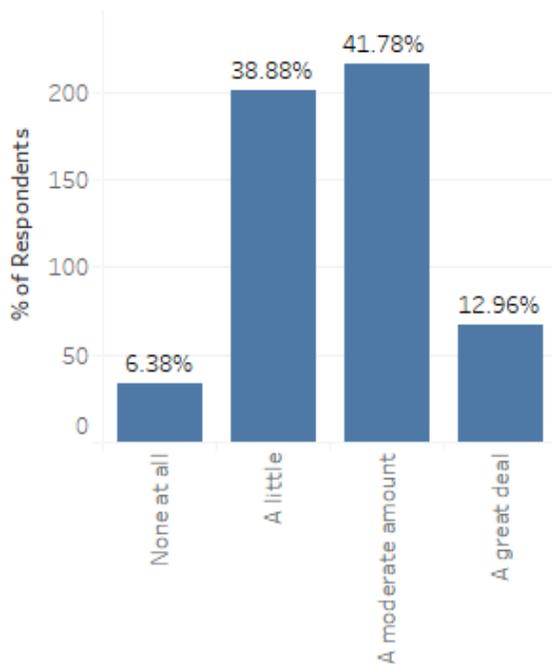


Figure 26. Results from Survey Question 20

As the researcher began to drill down to how disaggregated teacher cohorts answered Q20 via cross-tabulational data (Table 7, below), there were a few noticeable trends:

- Over 70% of all NBCT (71.87%) felt that the TPAS process did not engage a teacher in self-reflection and/or only facilitated self-reflective activities to a “small degree.”
- Close to 20% of all elementary school (ES) teachers (19.17%) responded that TPAS engaged teachers in self-reflection to a “great deal” and over 60% of all ES teachers (61.14%) reported that the TPAS process was self-reflective for a teacher to either a “moderate amount” and/or to a “great deal.”
- Over 50% of middle school (MS) teachers (53.22%), teachers that have been teaching for more than 20 years (53.66%), teachers with more than 20 formal TPAS ratings (55.55%), and high school (HS) teachers (51.82%) responded that the TPAS process was self-reflective for a teacher to either a “moderate amount” and/or to a “great deal.”

Table 7

Teachers' Responses to Survey Question 20 (in percent)

		Q20: How much of the TPAS process engages a teacher in self-reflection			
		None at all	A little	A moderate amount	A great deal
Different Demographic Cohorts	NBCT (n=36)	12.50%	59.37%	21.88%	6.25%
	Teaching more than 20 years (n=124)	8.94%	37.40%	44.72%	8.94%
	Teacher had more than 20 rated TPAS observations (n=61)	22.22%	22.22%	44.44%	11.11%
	Elementary School Teacher (n=193)	3.11%	35.75%	41.97%	19.17%
	Middle School Teacher (n=63)	9.68%	37.10%	46.77%	6.45%
	High School Teacher (n=110)	2.73%	45.45%	39.09%	12.73%
	Teachers rated on TPAS as "highly effective" (n=246)	5.35%	34.57%	44.03%	16.05%
	Overall	6.38%	38.88%	41.78%	13.96%

Q27 Impact of Self-Reflection

The researcher wanted to determine if larger percentages of teachers from different demographics self-reflect and, if so, what was the “impact” of such reflection on their self-reported development as a teacher? The first part of the analysis is given in Figure 27 below:

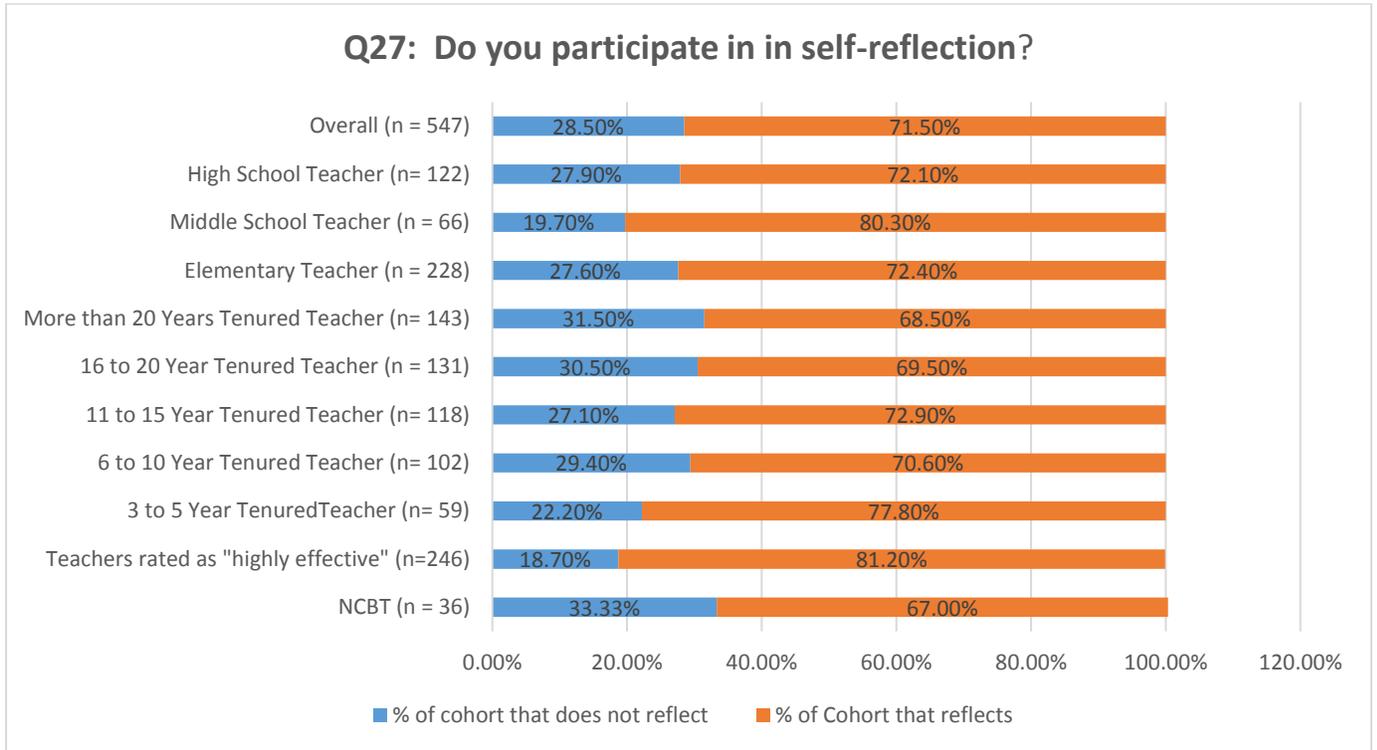


Figure 27. Results from Survey Question 27

One of the researcher’s initial hypotheses was that the NBCT Cohort would demonstrate a higher propensity to self-reflect – comparatively - than every other teacher cohort in this study; especially for those teachers that were *not* National Board Certified (i.e., non-NBCT). This was primarily due to the fact that to achieve National Board Certification (NBC), candidates had to go through a rigorous year of reflection in order to achieve such Certification. Hence, the researcher assumed that any NBCT would naturally participate in daily self-reflective activities, embedded as part and parcel of one’s (NBCT’s) daily routine. This propensity of NBCT to reflect would be evident in the following tangible capacities:

- The *degree of participation* in self-reflection
- A teacher's *professional belief in the impact* (effectiveness) of self-reflection; and
- The *frequency* with which teacher engage in self-reflective activities

Interestingly enough, the research did *not* confirm the initial hypothesis since it was demonstrated above that NBCT did not report the highest percentage of participants that engaged in self-reflection; rather, teachers that were rated as “highly effective” actually possessed the highest percentage of respondents who claimed that they self-reflect (i.e., 81.20%). Ironically, it was the NBCT Cohort that reported the *lowest* percentage of teacher cohorts that engage in self-reflection (i.e., 67.00%). Middle school teachers were the only other teacher demographic besides “highly effective” teachers that reported to have more than 80% of their cohort engage in self-reflection (i.e., 80.30%)

The second part of Q27 dealt with different teacher cohort's perception of the *impact* that self-reflection had on their development as a teacher. Teachers could respond to the stem of the question with either “a small amount”; “a moderate amount”; or a “great deal.” The results are much more in line with the researcher's hypothesis regarding NBCT and their propensity to reflect as over 40% of all NBCT (i.e., 41.30%) reported that they feel that reflection impacts their teacher effectiveness “a great deal.” Teachers that were rated as “highly effective” and all “3 to 5 year tenured teachers” were the cohorts with the next largest amount (i.e., 38.90%) of respondents that believe that reflection impacts their teaching.

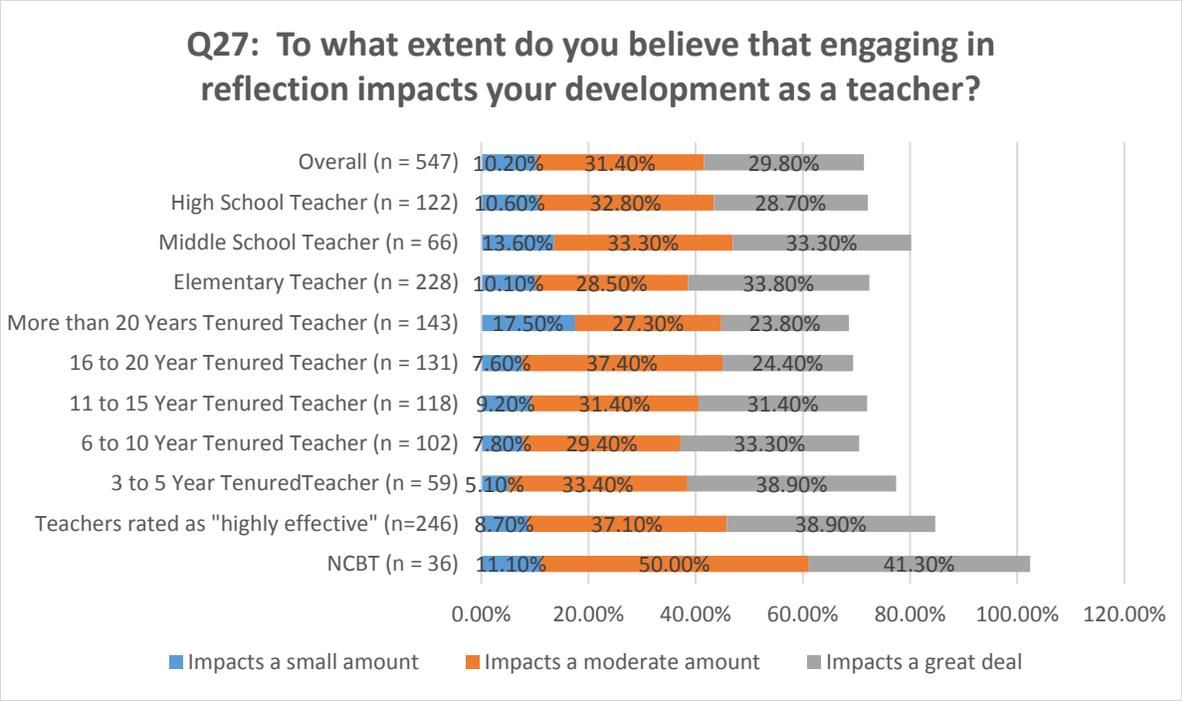


Figure 28. Q27 respondent results of different demographic groups regarding TPAS process

The next data collation in the figure (i.e., Figure 29) below represents the particular types of self-reflective activities in which teachers reported to have engaged. Overall, the largest percentage of responses consisted of teachers “audio-taping” their own teaching (16.297%); this was followed by teachers “reviewing student performance data” (15.678%); “completing a self-reflective check-list” (15.042%); and “engaging in unstructured reflection” (13.576%). Slightly less popular self-reflective activities were “video-taping oneself” (11.473%); “building a portfolio” (11.49%); and “journal writing” (9.306%). The least popular self-reflective activity reported by teachers was to “administer a student survey” to their class, with only 7.18% of all teachers participating in that particular type of reflection.

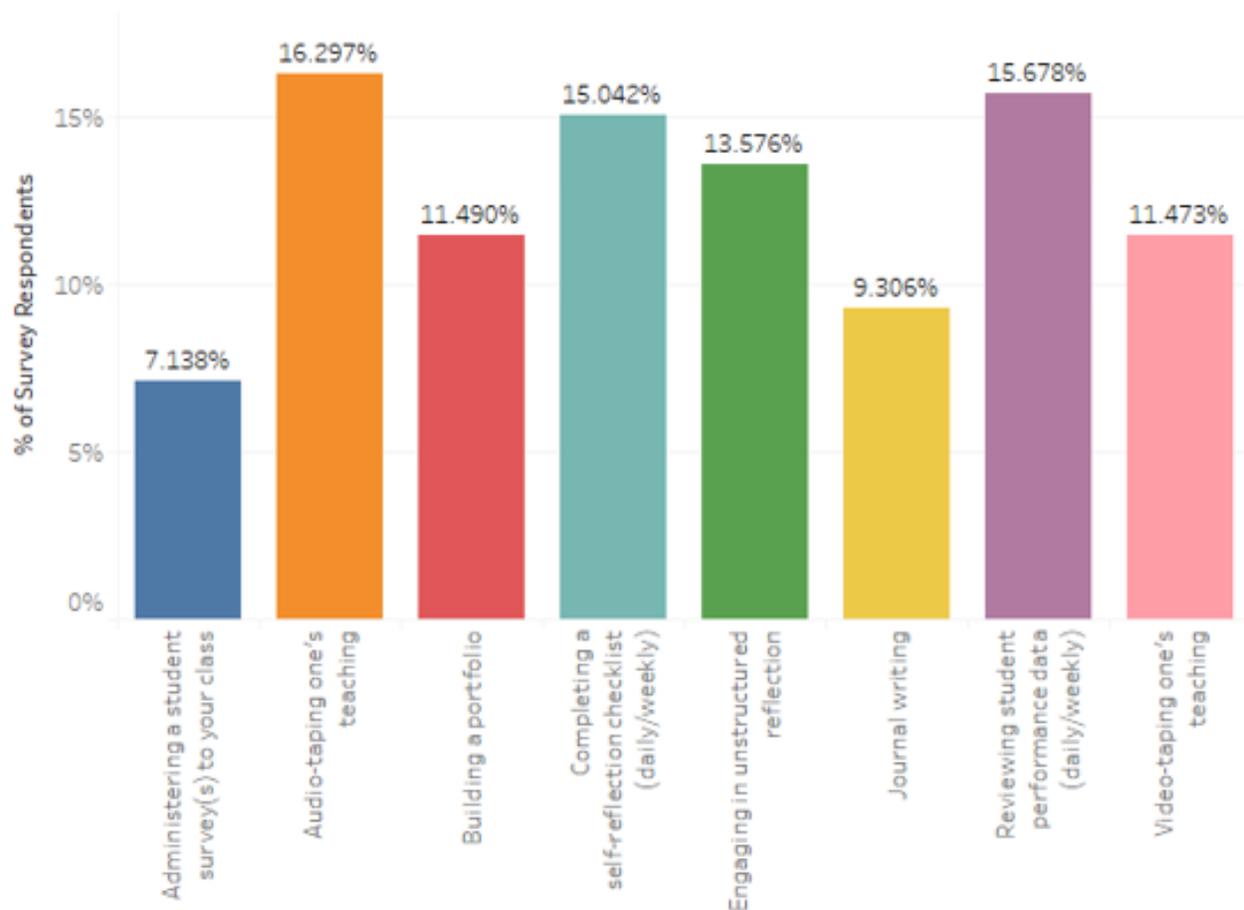


Figure 29. Percentage of teacher participation in self-reflective activities

Q28: Participation in Self-Reflection

The investigation then toggled over to Q28 and the researcher wanted to review the cross-tabulational data of two cohorts in particular: the NBCT cohort and the non-NBCT cohort to see if there would be substantive differences in the particular type of self-reflective activities in which each group would engage. While the data trend seemed somewhat consistent across each demographic, there were a few particular self-reflective activities that presented a considerable delta between the two cohorts. “Administering a student survey” “building a portfolio”, and “video-taping one’s class” were three specific reflective activities that possessed the greatest margin of difference between each cohort. Nevertheless, both the NBCT and the non-NBCT

groups exhibited somewhat similar responses when it came to “engaging in unstructured reflection”, “reviewing student performance data”, and “journal writing.”

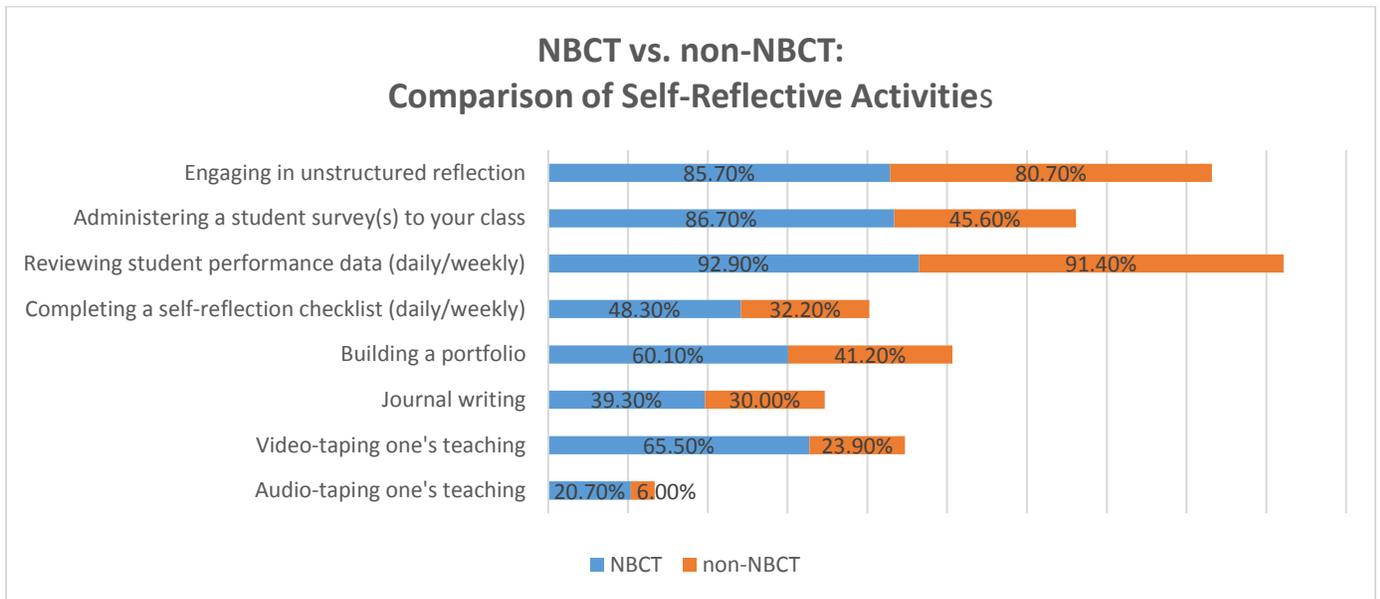


Figure 30. A comparison of self-reflective activities between NBCT and non-NBCT

The second part of Q28’s analysis dealt with the frequency with which each respective cohort (i.e., NBCT vs. non-NBCT) claimed to have either “moderately utilized” and/or “utilized self-reflective behaviors to a large degree”; the results are captured in Figure 31 below. Once again, the data is relatively surprising in that the researcher did not discern large gaps in the *frequency* of self-reflective activities between each cohort. In fact, the NBCT and the non-NBCT cohort have almost identical rates of reflective engagement, with over 90% of each cohort reporting that they engage in reflection - from a moderate to a large degree of frequency - for the two most popular self-reflective activities: “*engaging in unstructured reflection*” and “*reviewing student performance data.*” This data trend continued across other self-reflective activities as each cohort’s (NBCT and non-NBCT) self-reported degrees of frequency (from a moderate to a large degree) were relatively comparable for activities such as “*completing a self-reflection checklist*” (NBCT %: 86.7% vs. non-NBCT %: 87.0%); “*building a portfolio*” (NBCT %: 70.6%

vs. non-NBCT %: 69.5%); and “*journal writing*” (NBCT %: 72.7% vs. non-NBCT %: 62.9%), respectively. Both cohorts, equally, did not seem to engage in “*administer student surveys*” to their classes to any large degree (NBCT %: 46.2% vs. non-NBCT %: 49.6%).

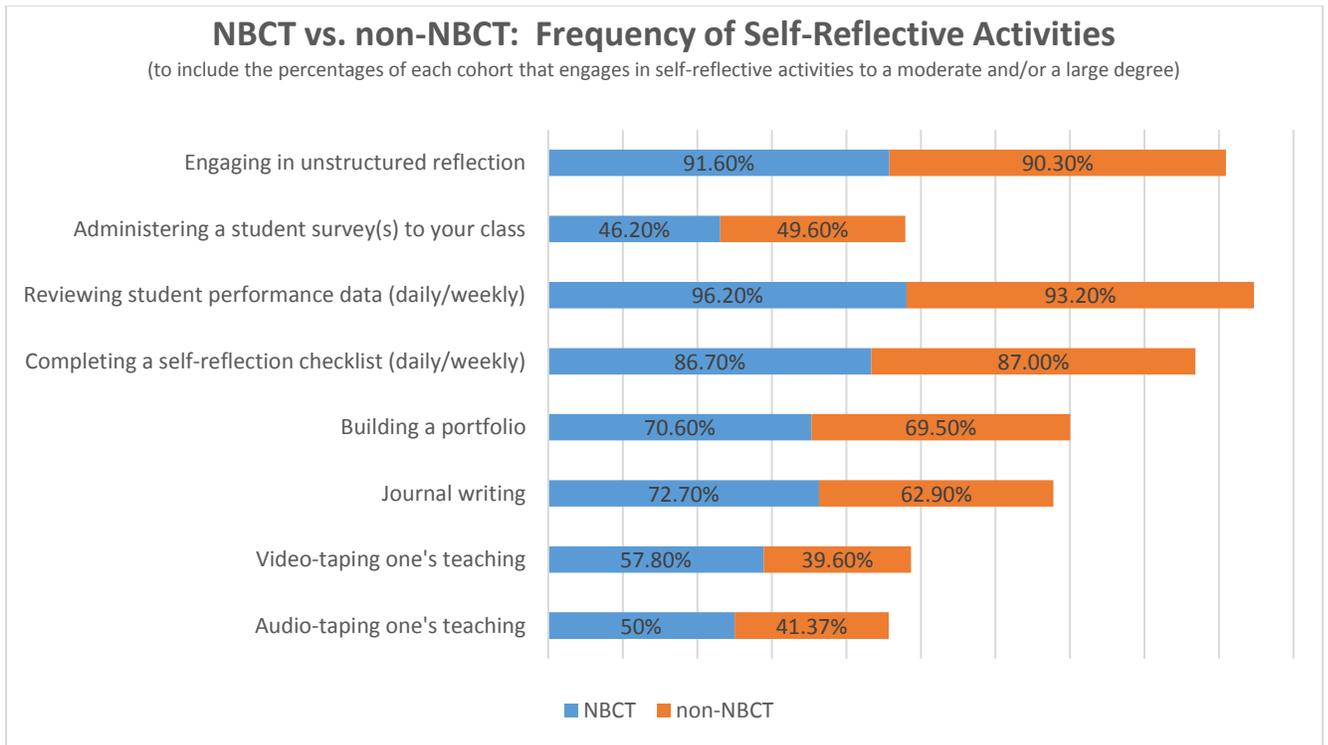


Figure 31. A comparison of the frequency of self-reflective activities between NBCT and non-NBCT

It should be noted that, on occasion, the non-NBCT cohort actually frequented more self-reflective activities than their NBCT peer cohort specifically when “*completing a self-reflection checklist*” (NBCT %: 86.7% vs. non-NBCT %: 87.06%) and “*administering student surveys*” (NBCT %: 46.2% vs. non-NBCT %: 49.6%). Nevertheless, the researcher was somewhat startled by the high degrees of self-reported frequency of the non-NBCT cohort in almost every different type of self-reflective activity; this was not at all expected.

Q29: Do Teachers know how to Self-Reflect?"

As the figure (i.e., Figure 32) below denotes, an overwhelming majority of teachers (85.37%) self-proclaim that they know how to “self-reflect”...and do so often. Approximately 12% of survey respondents insist that they know how to “self-reflect” but choose not to since (they) “do not have the time.” Less than 1% of teachers self-report that they do not know how to self-reflect and/or they “know how to self-reflect” but do not believe that such actions assist instructional delivery.

Q29: As a teacher, I generally know how to “self-reflect.”

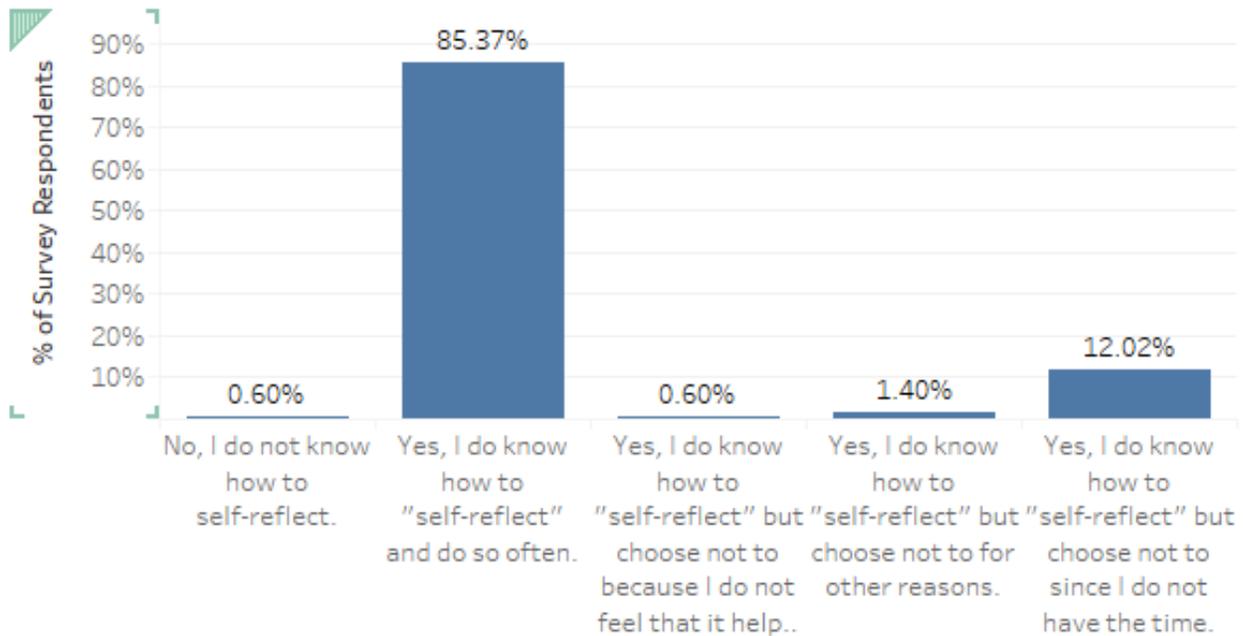


Figure 32. Results from Survey Question 29

What is especially acute is the Q29 breakdown of the demographic data in another cross-tabulation, 94.12% of all “highly effective” teachers responded with the highest percentage of teachers that self-reported that not only do they know how to self-reflect but they do so often. NBCT reported the next highest percentage of their cohort that actively self-reflects, with over

93% of them claiming that they do so. All other teacher cohorts (teachers with more than 20 years of experience; teachers with more than 20 observations; elementary school teachers; and high school teachers) all responded that over 80% of their cohort knows how to self-reflect and does so often. Middle school teachers were the only cohort that did not hit the 80% threshold of self-reported *self-reflection* (77.42%) and they also had the highest number of teachers' claim that they did not know how to self-reflect.

Table 8

Teachers' Responses to Survey Question 29 (in percent)

		Q29: As a teacher, I generally know how to “self-reflect.”				
		No; I do not know how to self-reflect	Yes; I do know how to self-reflect and do so often	Yes; I do know how to self-reflect and but choose not to because I do not feel that it helps	Yes; I do know how to self-reflect but choose not to for other reasons	Yes; I do know how to self-reflect but choose not to since I do not have the time
Different Demographic Cohorts	NBCT (n=36)	0%	93.75%	0.00%	0.00%	6.25%
	Teaching more than 20 years (n=124)	0.81%	82.93%	0.81%	2.44%	13.01%
	Teacher had more than 20 rated TPAS observations (n=61)	0.00%	85.00%	0.00%	5.00%	10.00%
	Elementary School Teacher (n=193)	0.00%	88.30%	0.00%	1.60%	10.11%
	Middle School Teacher (n=63)	3.23%	77.42%	0.00%	0.00%	19.35%
	High School Teacher (n=110)	0.96%	82.69%	0.00%	0.96%	15.38%
	Teachers rated on TPAS as “highly effective” (n=246)	0.00%	94.12%	0.00%	0.00%	5.88%
	Overall	0.60%	85.37%	0.60%	1.40%	12.02%

Research Question Four: Factors that Impact Teacher Effectiveness

The data related to Research Question Four came from SMCPs teacher respondents to Survey Questions 13, 24, 25, and assorted demographic data, respectively.

Research Question	Survey Questions
4. Do tenured teachers that have been evaluated in the Teacher Performance Assessment System (TPAS) grow in teacher effectiveness over time (as measured by TPAS ratings) and, if so, what factors account for this growth?	<ul style="list-style-type: none"> • Assorted demographic data such as Q1(years of teaching); Q3(NBCT or not); Q5(highest degree conferred); Q6(placement level of teacher); Q12 (total number of TPAS observations in career); • Q13(first TPAS rating); Q26(last TPAS rating) • Q24. Which of the following professional development activities has positively impacted your instruction the most? • Q25. As a teacher, how do you see yourself spending most of your "learning time" in the schoolhouse? • Takeaways from Research Question #3

Figure 33. Crosswalk of Research Question Four with Survey Questions

The researcher wanted to specifically address whether or not teachers that have been evaluated in TPAS grow in teacher effectiveness over time (as measured by TPAS ratings) and what factors contribute to this growth? The implication here is that TPAS is not only a tool – like most teacher evaluation metrics - for teacher compliance and quality control but as a *formative* mechanism that builds teacher effectiveness over time; at least that is the hope.

Given that hypothesis, it would seem logical to this researcher that if an evaluation system was, in fact, *formative*, then the more frequent a teacher is evaluated within said system, their performance on their final evaluation would be optimized. That is, there would be a strong, positive correlation that the more times an individual teacher was observed/evaluated with the TPAS Framework, the closer one would achieve the highest TPAS rating of “highly effective.” In turn, this conclusion would seamlessly liaison with Roger Thorpe’s query of “*How does a teacher evaluation metric maximize the chances that those who remain in the teaching profession become accomplished practitioners?*” Hence, Research Question 4, will initially

unpack the results of the relationship between the total number of observations and a teacher's last, self-reported TPAS rating to see if there are any identifiable patterns. What should also be noted is that some of the analyses from Research Question 4 will assist the researcher in addressing some of the issues raised from Research Question 3.

The Total Number of Observations and Last TPAS Rating

As we review the figures below, there does not seem to be any substantive correlation between the number of formal observations a teacher experiences and their corresponding TPAS rating. The researcher incorporating Q12 and Q26 from the survey and performed a cross-tabulation of the respondents. It should be noted that any average TPAS rating over 3.5 is considered to be "highly effective" as we translate discrete, integer evaluation ratings. If TPAS is truly formative, what the data should present is that as teacher undergo more formal TPAS observations, their average TPAS rated should summarily increase as well; this is clearly not the case. As evidenced, tenured teachers with 1 to 2 formal TPAS observations have reported that their average TPAS rating (out of a perfect 4.0) is 3.5588; an average that puts those teachers in the "highly effective" category. Comparatively, if we review the 20 teachers with more than 20 observations to their name, their average TPAS rating is marginally smaller, with an average of 3.5101. Nevertheless, those 228 teachers that have had 6 to 10 observations over their career (TPAS rating of 3.4561) or teachers in that 11 to 15 observations' category (TPAS rating of 3.3134) have significantly lower average TPAS ratings despite being observed more frequently. See Table 9 and Figure 34 below for more detailed respondent data.

Table 9
Number of Observations and Corresponding TPAS Ratings

Number of Observations and TPAS Ratings

RQ#4 - Do more observations lead to higher TPAS ratings?	% of Survey Respondents (total)	Average TPAS Rating (out of 4.00)
1 to 2 observations	6.8% (34)	3.5588
3 to 5 observations	23.7% (117)	3.4103
6 to 10 observations	44.9% (228)	3.4561
11 to 15 observations	13.2% (67)	3.3134
16 to 20 observations	8.3% (42)	3.5476
More than 20 observations	3.9% (20)	3.5101

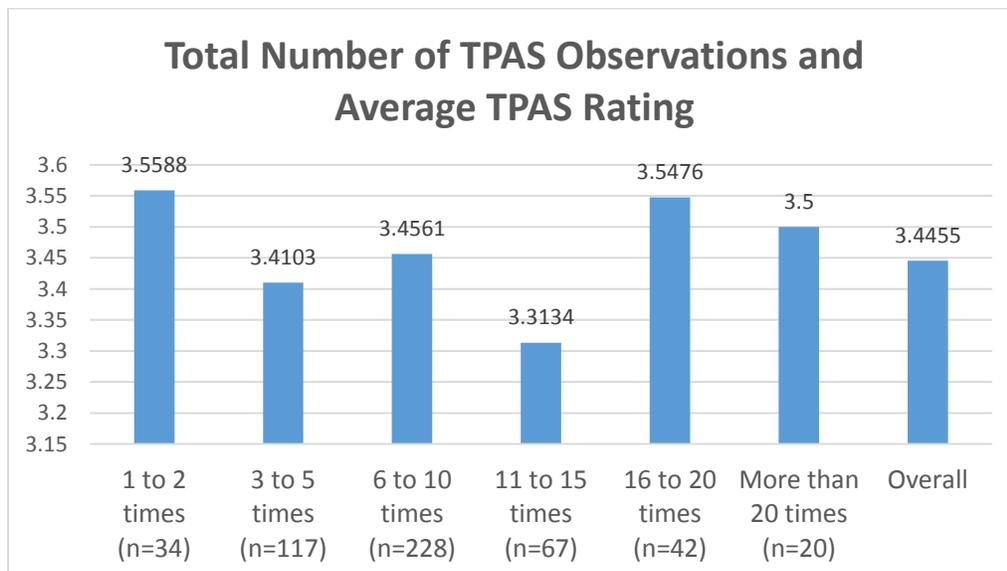


Figure 34. Total number of TPAS observations and average TPAS rating

Years of Service and Last TPAS Rating

The experience level of the teacher was another factor that the researcher wanted to consider when discussing teacher effectiveness. That is, do teachers grow in effectiveness by the sheer fact that they are a veteran teacher? In essence, one's average TPAS rating should appreciate as the years of service a teacher puts into the system summarily increases. As the figure (i.e., Figure 35) below denotes, there is nothing definitive in the data that validates the aforementioned hypotheses that as years of service increase; so, too, does one TPAS rating. If that was the case, then those teachers with more than 20 years of service would have the highest average TPAS rating.

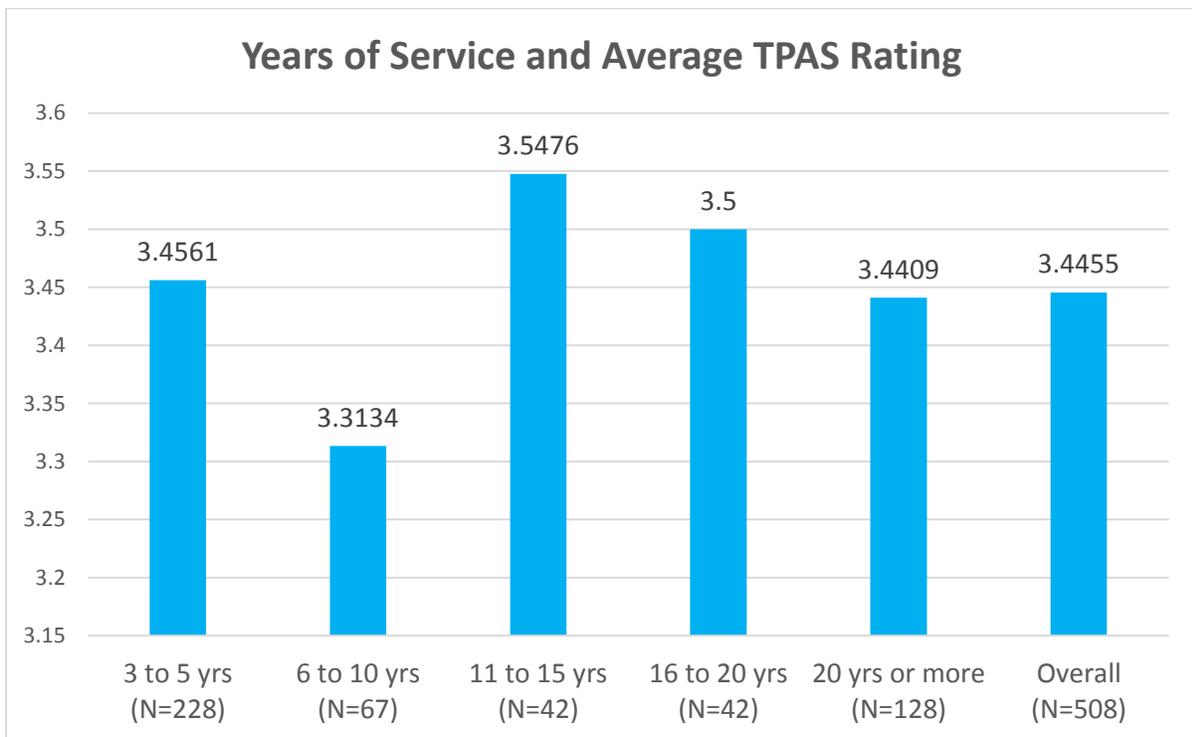


Figure 35. Years of service and average TPAS rating

NBCT and the Last TPAS Rating

As the researcher continues to evaluate factors that build teacher effectiveness in staff, it is clearly demonstrated in Figure 36 below that the NBCT cohort has a significantly higher average TPAS rating (3.5833) as compared to the non-NBCT cohort (3.4352) and the overall average TPAS rating (3.4455), respectively.

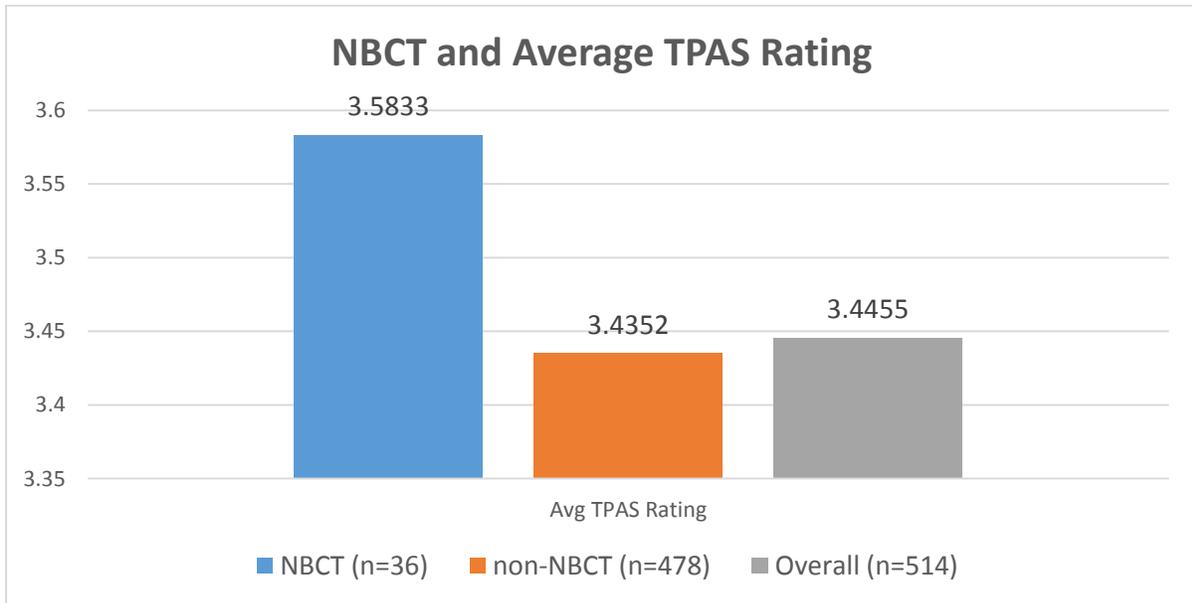


Figure 36. NBCT and average TPAS rating

Frequency of Unstructured Reflection and the Last TPAS rating

The study continues to evaluate factors that impact teacher quality and effectiveness over time. The next cohort of teachers to be scrutinized are those teachers who admittedly engage in unstructured reflection to either a moderate and/or a high degree. The cross tabulational results from Q26 and Q28 are captured in the table below. The researcher, in this first calculation, wanted to parse out degrees of reflection and attempt to quantify, in a general sense, the effect of actively “self-reflecting” (at least more than a moderate degree) in one’s average TPAS rating.

Table 10

Frequency of Teacher Reflection and Average TPAS Ratings (in general)

Frequency of Reflection and Average TPAS Rating (in general)

Q28: Participation and Frequency of Unstructured Self-Reflection	% of Survey Respondents (total)	Average TPAS Rating (out of 4.00)
<i>Yes; I generally engage in unstructured reflection to a moderate and/or a high degree.</i>	80.7% (393)	3.4461
<i>No; I do not engage in unstructured reflection to a moderate and/or a high degree.</i>	19.3% (94)	3.3804

Not only was the researcher struck by the high number of teacher respondents who answered affirmatively that they admittedly self-reflect at least to a *moderate* degree (i.e., 80.7%), the data also speaks to the fact that one’s average TPAS rating is higher as well, albeit by only about 2%; which was somewhat disappointing. More analysis would be needed to see if this results is statistical significant.

Since there was a marginal impact in average TPAS rating when teachers engage in “self-reflection” to at least a moderate degree, the researcher wanted drill down even further and inspect whether or not self-reporting teachers that claim to self-reflect to a large degree had a more substantial, higher average TPAS ratings than their peers that choose not to “self-reflect” as much. In Table 11 and Figure 37 below, there does seem to be an impact in average TPAS ratings for teachers that self-reflect to a large degree as evidenced by that cohort of teachers securing a 3.586 average TPAS rating, which was the highest recorded average TPAS rating amongst all demographic cohorts in this study.

Table 11

Differentiated Frequency of Teacher Reflection and Average TPAS Ratings

Q28: Participation and Frequency of Self-Reflection	# of Survey Respondents	% of Survey Respondents	Average TPAS Rating (out of 4.00)
<i>I do not self-reflect</i>	66	13.8%	3.412
<i>I self-reflect to a small degree</i>	60	12.5%	3.281
<i>I self-reflect to a moderate degree</i>	184	38.3%	3.482
<i>I self-reflect to a large degree</i>	170	35.4%	3.586
Overall Totals	480	100.0%	3.4455

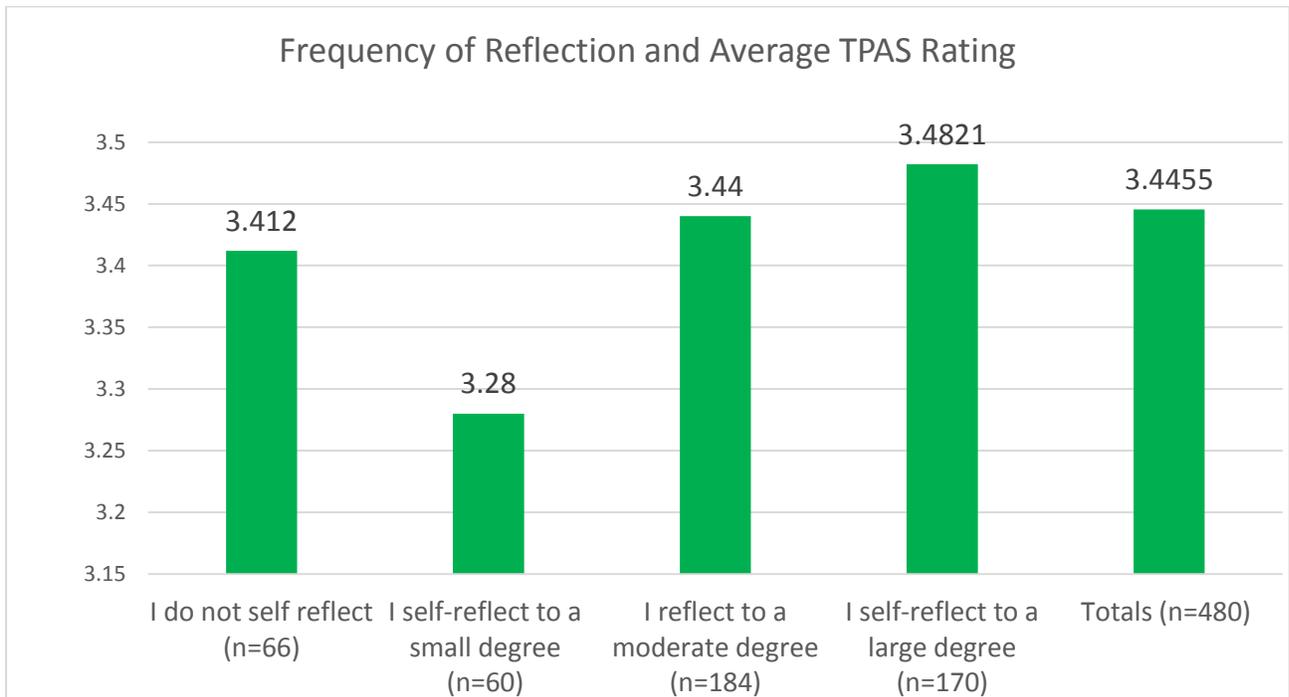


Figure 37. Frequency of reflection and average TPAS rating

Frequency of Reflection, NBCT and the Last TPAS Rating

Questions 3, 26, and 28 on the teacher survey addressed the cross tabulation of frequency of reflection, NBCT Certification, and a teacher’s most recent TPAS rating. The initial reaction from Table 12 below is that non-NBCT self-reflect more frequently *and* to a larger degree than NBCT. Close to 30% of all non-NBCT respondents reported to engage in self-reflection to a “large amount” whereas NBCT responded that only about 20% of all NBCT reflect to a “large amount.” Notwithstanding, over 60% percent of all non-NBCT reported that they self-reflect to either a “moderate” and/or to a “large amount” which is more than the 55% of NBCT over that same metric. Similarly surprising to the researcher was that one out of every three NBCT survey respondents claim to not self-reflect at all; this result is quite shocking given what a teacher must go through to achieve their National Board. Perhaps the National Board experience does not build the requisite capacity in a teacher to self-reflect like what was initially hypothesized by the researcher. In any case, these results will help us to further augment issues raised from Research Question 3 regarding the extent to which teachers “self-reflect.” Taken together with the previous section, the researcher can conclude that what may be more telling is the *degree to which teachers self-reflect*, as evidenced in the following table (i.e., Table 12).

Table 12

Frequency of NBCT Reflection and Average TPAS Ratings

Q28: Participation and Frequency of Unstructured Self-Reflection	% of Survey Respondents (total)	None	Small Amount	Moderate Amount	Large Amount	Average TPAS Rating (out of 4.00)
<i>NBCT</i>	6.2% (36)	33.3% (12)	11.1% (4)	36.1% (13)	19.4% (7)	3.582
<i>non-NBCT</i>	93.8% (547)	28.5% (156)	10.2% (56)	31.4% (172)	29.8% (163)	3.431

Given the previous results and analyses, the researcher wanted to drill down further with the NBCT Cohort to see if, in fact, degrees of reflection within this group would have any impact on average TPAS rating. It was confirmed that as NBCT who admittedly self-reflect to a *large degree*, their average TPAS rating was considerably higher (3.583 TPAS rating) than their peer NBCT group (3.25 TPAS rating) who reported that they self-reflected at an amount *less than a large degree* (which could have been either no reflection at all, reflection to a small and/or a moderate degree). Once again, it should be noted that degrees to which teachers self-reflect impact teacher evaluations more than what was previously thought; in this particular case, there was a positive effect of approximately 10.3% on a local teacher evaluation system as a result of more teachers becoming more intense reflective practitioners. In other words, it may not matter much that teachers *participate* in self-reflection; however, what does seem to influence teacher effectiveness over time is *the degree to which teachers engage in self-reflection*; at least for teachers that have secured their National Board.

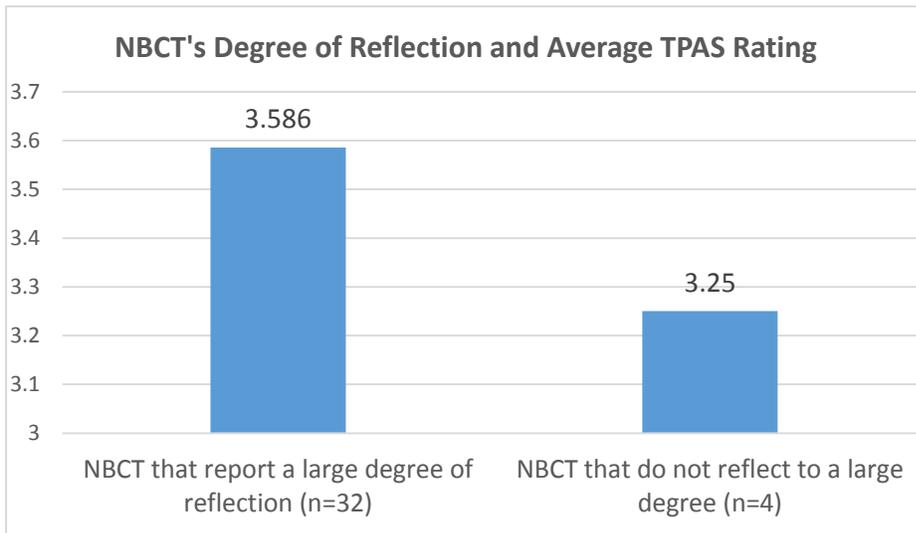


Figure 38. NBCT’s degree of reflection and average TPAS rating

Frequency of Reflection, Experience Level and the Last TPAS Rating

Questions 1, 26, and 28 on the teacher survey addressed the cross tabulation of frequency of reflection, experience level and a teacher’s most recent TPAS rating. In the previous section, it was shown that the degree to which a National Board teacher reflects substantively impacts one’s TPAS rating. For the experienced teachers’ cohort, there seems to be an inverse relationship between *more experienced* teachers and the degree to which they reflect; that is, as teachers become more experienced, respondents reported that they engage in less amounts of unstructured self-reflection. In Table_ below, close to 40% of teachers with three to five years of experience (38.9%) reported to have engaged in the largest amounts of self-reflection while only around 20% of teachers with more than 20 years of experience (23.8%) self-reflect in kind; which was the lowest reported teacher cohort - experience level - in this survey. Notwithstanding, over 30% of the most experienced teachers in this survey (teachers with more than 20 years of experience) responded that they do not self-reflect at all (31.5%).

Table 13

Years of Experience, Frequency of Reflection and Average TPAS Ratings

Q28: Participation and Frequency of Unstructured Self-Reflection	% of Survey Respondents (total)	None	Small Amount	Moderate Amount	Large Amount	Average TPAS Rating (out of 4.00)
<i>3 to 5 years of teaching</i>	10.7% (59)	22.1% (13)	5.1% (3)	33.4% (20)	38.9% (23)	3.4038
<i>6 to 10 years of teaching</i>	18.4% (102)	29.4% (30)	7.8% (8)	29.4% (30)	33.3% (34)	3.3809
<i>11 to 15 years of teaching</i>	21.3% (118)	27.1% (32)	9.2% (12)	31.4% (37)	31.4% (37)	3.4953
<i>16 to 20years of teaching</i>	23.7% (131)	30.5% (40)	7.6% (10)	37.4% (49)	24.4% (32)	3.5083
<i>More than 20 years of teaching</i>	25.9% (143)	31.5% (45)	17.5% (25)	27.3% (39)	23.8% (34)	3.4409

Nevertheless, teaching experience seems to mute any lack of reflection in a teacher's TPAS rating as evidenced by the appreciation of TPAS ratings commensurate with years of experience over time: as teaching experience increases, the degree to which teachers self-reflect in large amounts decreases while the average TPAS ratings increase over time. This appreciable data trend in average TPAS ratings held true for all levels of teaching experience but one: teachers with the most experience/more than 20 years of service. Teachers in the 16 to 20 year band of experience demonstrated the highest average TPAS rating out of every experience level cohort in the survey with a rating that produce an overall "highly effective" level (3.5083) for all teachers in this cohort.

Frequency of Reflection, Leveled Teacher and the Last TPAS Rating

Questions 6, 26, and 28 on the teacher survey addressed the cross tabulation of frequency of reflection, leveled teacher, and a teacher's most recent TPAS rating. The researcher also wanted to review the level at which teachers are placed (i.e., elementary, middle, or high school) to see if either participating in reflection and/or the degree to which teachers self-reflect impact their teaching effectiveness as represented by their TPAS rating. In Table_ below, the elementary school (ES) teacher not only possessed the highest TPAS rating but also engaged in the largest amount of self-reflection, with close to 35% of all ES teachers reporting that degree of reflection.

Table 14

Leveled Teacher, Frequency of Reflection and Average TPAS Ratings

Q28: Participation and Frequency of Unstructured Self-Reflection	% of Survey Respondents (total)	None	Small Amount	Moderate Amount	Large Amount	Average TPAS Rating (out of 4.00)
<i>Elementary Teacher</i>	53.0% (228)	27.6% (63)	10.1% (23)	28.5% (65)	33.8% (77)	3.474
<i>Middle School Teacher</i>	16.9% (66)	19.7% (13)	13.6% (9)	33.3% (22)	33.3% (22)	3.338
<i>High School Teacher</i>	30.1% (122)	27.9% (34)	10.6% (13)	32.8% (40)	28.7% (35)	3.435

Despite two out of all three middle school (MS) teachers (66.7%) claiming to have engaged in at least a moderate amount of self-reflection, their average TPAS rating was the lowest (3.338) of all leveled cohorts in the survey. High school (HS) teachers reported to that they engage in the lowest amounts of reflection, with approximately 28% of them (27.9%) stating that they do not engage in any self-reflection coupled with 28.7% of HS teacher respondents reporting that they engage in large amounts; the highest and lowest reporting percentages for leveled teachers in this survey, respectively. Next steps for research would be to attempt to control the inter-rater reliability at each level and/or parse out the intra-degree to which each leveled cohort engages in reflection commensurate with their average TPAS rating (i.e., the average TPAS rating of ES teachers that self-reflect in large amounts compared to the average TPAS ratings of all other ES teachers).

Specific Types of Professional Development for Building Teacher Effectiveness

Teachers were surveyed on the types of professional development that have been most beneficial in impacting their instruction. Teachers were given nine categories (including an “other” option) in which respondents were asked to identify which particular professional development activities has most positively impacted their instruction and, subsequently, their teacher effectiveness. These responses are to be used to help the researcher work with the local school system to help shape future professional development workshops in various areas that teachers felt were most beneficial in building their teacher quality. Responses were also collated by three demographic cohorts of teachers: NBCT; non-NBCT, and those teachers rated as “highly effective.” The “*TPAS process and lesson observations*” category was an appropriate option for teachers given the proclaimed “formative” nature of TPAS and the nature of this study. Interestingly enough, respondents were none too fond of this particular professional development as being impactful in building their teaching capacity as close to 1% of teachers in both the NBCT (0.90%) and the “highly effective” (1.0%) cohorts, respectively, similarly reported their sentiments.

For “highly effective” teachers, the highest percentage of respondents of most impactful professional development consisted of “*professional learning communities*” (21.90%); “*peer coaching and mentoring*” (20.0%); and, to a lesser degree, “*formal, credit-bearing coursework at a higher institution of learning*” (13.0%). “*Conferences*” (5.0%) and “*informal, continuing professional credit coursework*” (7.0%) were two of the lowest rated categories of professional development as reported by “*highly effective*” teachers (the “*TPAS process and lesson observation*” category was rated lowest by “*highly effective*” teachers).

For the NBCT's cohort, the "*other*" category (21.90%) was rated by this group as most impactful in building their instructional prowess, followed by the act of "engaging in self-reflection" (18.80%). What NBCT cited as "other" professional development, their write-in responses were as follows:

- "Informal shares with colleagues"
- "National Board Certified process" (6 times)
- "Time to plan individually and formal time set aside during the duty day to team plan"
- "Peer to peer walkthroughs"

"*Online, self-guided professional development*" (0.00%) and the "*TPAS process and lesson observations*" (0.90%) were the lowest rated professional development activities by NBCT.

For all other teachers, there was a non-NBCT category (not including either NBCT or "highly effective" rated teachers) for the researcher to analyze. Similar to NBCT and the "highly effective" teachers cohort, non-NBCT responded that they were positively impacted by "professional learning communities" (21.90%) and "peer coaching and mentoring" (20.70%); these responses were consistent with the two aforementioned cohorts. Somewhat out of the ordinary given the data heretofore was that 21.0% of all non-NBCT felt that the "*TPAS process and lesson observations*" were impactful and helped to build their teacher effectiveness; this result was in direct contrast to the survey responses fielded by the NBCT and the "highly effective" teachers' cohorts.

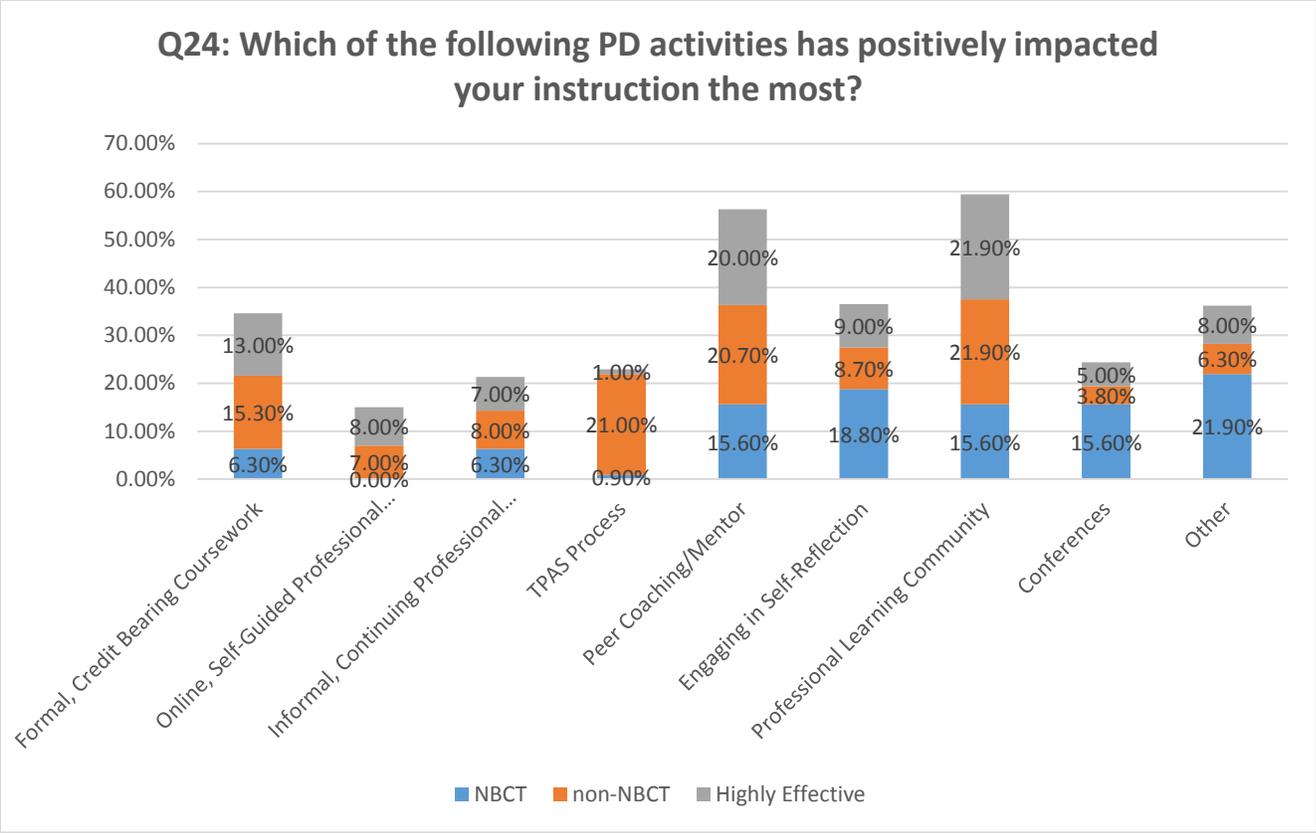


Figure 39. Results from Survey Question 24

Conclusion – Resolution of the Research Questions

If the quest for any school system is to strive for all of its teachers to be rated as “highly effective”, then it would seem wise to synthesize a comprehensive data profile of these teachers who have already been rated as “highly effective” in a teacher evaluation system and work towards building similar capacities in those teachers *without* that rating. Throughout this investigation, the researcher has consistently reviewed various demographic data to see if specific data trends and identifiable patterns would present themselves in tangible ways. Looking at National Board Certification; “highly effective” rated teachers; experience level; highest degree conferred; the number of total career observations a teacher has experienced; and, even, leveled teacher placement, this researcher had tried to unpack in every cross tabulation a snippet of analytical dependence that would shed more light on these issues.

The researcher formulated conclusions to the four primary research questions as they related to the literature review and the analyzed data. **Research Question One** (RQ1) queried tenured teachers about their perception that the current teacher observation system (TPAS) impacts the quality of their instruction. The researcher asked teachers what was the overall impact of TPAS on their instructional prowess over time and which part of the TPAS process had been most impactful to that end? Also, was there a disconnect from what teachers experience with the current TPAS Framework as “it presently is “and what this TPAS experience “could be” in the future? Furthermore, the researcher wanted to quantify teachers’ opinion of TPAS as it related to providing assistance for teachers in completing their professional duties. The following abbreviated survey questions were utilized in answering Research Question 1(RQ1):

- Q17: TPAS as a tool for Teacher Evaluation (cohorts)?

- Q18: What is overall impact of TPAS?
- Q19: Which part of the TPAS process is most impactful?
- Q21: Perceptions of TPAS up to this point?
- Q22: Perceptions of what TPAS could be?
- Q23: Perception of TPAS with respect to professional duties?

The results for RQ1 are below:

- Most teacher cohorts rated TPAS somewhere in the middle of the continuum, with the majority of all respondents claiming that they were either *somewhat dissatisfied*, *neutral*, or *somewhat satisfied* with TPAS and its ability to help build a teacher's overall capacity.
- The cohorts with the strongest reaction *against* the TPAS process and responded that they were *very dissatisfied* was evidenced in two particular demographics: teachers that had experienced more than 20 rated observations and high school teachers.
- It should be noted that among that same demographic of teachers with more than 20 rated TPAS observations, close to 9% of these same teachers (8.99%) were *very satisfied* with TPAS and its ability to build overall capacity. Hence, we can conclude that TPAS elicits the strongest sentiment – both *very satisfied* and *very dissatisfied* – with those teachers who have been formally observed the most with TPAS.
- By omitting all *neutral* responses from all disaggregated teacher cohorts and by combining responses that were either *very dissatisfied* or *somewhat dissatisfied* together coupled with combining those responses that were either *very satisfied* or *somewhat satisfied* in the same vein, the overall TPAS perception of all teacher cohorts seems to be *very or somewhat satisfied* with TPAS' ability to build overall teacher capacity; close to 70%

- Close to 70% of all survey respondents (i.e., 69.84%) perceive TPAS to be more than a *quality control* tool for teachers and that TPAS encompasses a *compliance* factor as well.
- Over 40% of all teachers (42.11%) believe that TPAS is not only an instrument for compliance and teacher quality control but also one that assists in the professional development as well for teachers.
- Over 20% of all teachers (i.e., 21.27%) self-reported that TPAS had no impact on their development as a classroom teachers and an additional 43.54% of teachers felt that TPAS had only a small impact on their performance. All told, close to 65% of all respondents (i.e., 64.81%) characterized the impact of TPAS on their own teacher effectiveness as either none and/or minimal across all demographics no outliers
- The one outlier cohort that reported the largest impact of TPAS (i.e., 16.67%) on their teaching were those teachers that had experienced more than 20 observations in their career; this response was more than 10 percentage points above the overall survey response on Q18 for TPAS having a *large impact* on teaching development (i.e., 5.37%).
- Teachers overwhelmingly responded (i.e., 67.29%) that the *post-conference* was the most impactful part of the TPAS process on one's teaching prowess; it was not the *actual observation* (9.35%) nor the *lesson-planning* (17.76%), respectively. *Lesson planning* was the 2nd most impactful; the least impactful part of the TPAS process was the *pre-conference* per survey respondents.
- Survey respondents were asked to respond to differentiated teaching duties and record the degree to which each respondent was “*satisfied*” or “*dissatisfied*” (including “*very satisfied*” or “*very dissatisfied*”) with TPAS and its ability to build capacity in teachers. Respondents were most positive with TPAS' ability to “*engage teachers in self-*

reflection”; “*lesson planning and using curricular resources*”; “*differentiating instruction*”; and “*implementing the Common Core.*”

- Teachers did not rate TPAS as being particularly effective in helping teachers deal with the “day to day” work flow of the job and “classroom management.”
- Over 46% of all teachers (46.03%) most often reported that past TPAS experiences facilitated a “*reflection*” of some capacity with regards to their teaching. Only 1.55% of all teachers responded that their past experiences with TPAS were energizing. Other rather low respondent marks for past TPAS experiences included TPAS being characterized as impactful (2.71%); and/or having a lack of *engagement* (5.03%).
- The greatest divide of TPAS characterizations resided with the notion that TPAS was “*supportive*”, encompassing an absolute delta value of 22.5% from all respondents; the next largest delta value for teachers was TPAS being characterized as “*impactful*” ($\Delta = 21.6\%$) between *past* experiences ($\Delta=2.7\%$) of TPAS with what TPAS *could be* in the future ($\Delta=24.3\%$). “*Poor use of time*” ($\Delta = 18.3\%$), and TPAS being “*reflective in nature*” ($\Delta = 16.3\%$) represented the next two greatest divides for teacher respondents. The smallest delta value for survey respondents resided with the TPAS characterization that it “*lacked engagement*” ($\Delta = 5.1\%$), implying that present TPAS implementation is closest to this reality and that the local evaluation system actually is “*engaging*” for teachers.

Research Question Two (RQ2) investigated which of the four component Domains of TPAS most effectively improve a teacher’s effectiveness over time. For these survey questions, there were over 500 respondents ($n=501$). The following abbreviated survey questions were utilized in answering Research Question 2 (RQ2):

- Q14. Which singular TPAS Domain most directly impacts your teaching skill?
- Q15. Which singular TPAS Domain is least likely to impact your teaching skill?
- Q16. Rank-order each TPAS Domain that impacts your teaching skill from least effective to most effective?

The results for RQ2 are below:

- Over 50% of all teachers (i.e., 54.46%) self-reported that Domain 3 – Instruction – was the most impactful TPAS Domain with regards to improving their own teacher effectiveness. In reviewing the cross-tabulated demographic data, some teacher cohorts reported higher levels of confidence in Domain 3 and its impact on teacher quality. That is, close to 60% of elementary school teachers, teachers that have taught for more than 20 years, and the NBCT’s cohort all responded higher degrees of confidence in TPAS Domain 3 than their peers. Only 43% of all middle school teachers reported that the *Instruction* Domain was most important in improving their teaching skill.
- Interestingly enough, 21.71% of respondents – overall - equally felt that the *Planning and Preparation* (Domain 1) and *The Learning Environment* (Domain 2), respectively, were the next most impactful TPAS Domains.
- Over 84% of all survey respondents (84.11%) reported that the least impactful TPAS Domain was Domain 4: Professional Responsibilities.
- NBCT was the only demographic cohort that rank ordered Domain 2 - *The Learning Environment* – as the next most impactful Domain after *Instruction*.

Research Question Three (RQ3) addressed the extent to which tenured teachers engage in self-reflection about their teaching and which particular cohorts have more of a propensity to “self-reflect” given a demographics’ degree of participation in self-reflection and their belief that

such reflection impacts teacher quality. Not to mention, the researcher also wanted to document from a teacher's vantage point how much embedded reflection exists in the TPAS process and, once again, the research was to be disaggregated down to various cohorts such as leveled teachers; years of teaching experience, the total number of TPAS observations; NBCT; and those teachers rated as "highly effective." The following abbreviated survey questions were utilized in answering Research Question 3 (RQ3):

- Q20: Teacher perception of self-reflection embedded within the TPAS Process
- Q27 Quantify the impact of self-reflection in differentiated teacher cohorts.
- Q28: Quantify a teacher's participation in self-reflection in differentiated teacher cohorts.
- Q28: Quantify the frequency of self-reflection in differentiated teacher cohorts.
- Q29: As a teacher, do you know how to self-reflect?

The results to RQ3 are below:

- Most teachers – regardless of demographic – reported that TPAS embeds a significant amount of self-reflection in its process already;
- Over 70% of all NBCT (71.87%) felt that the TPAS process did not engage a teacher in self-reflection and/or only facilitated self-reflective activities to a "small degree."
- Close to 20% of all elementary school (ES) teachers (19.17%) responded that TPAS engaged teachers in self-reflection to a "great deal" and close to 60% of all ES teachers reported that the TPAS process was self-reflective for a teacher to either a "moderate amount" and/or to a "great deal."
- Over 50% of middle school (MS) teachers (53.22%); teachers that have been teaching for more than 20 years (53.66%); teachers with more than 20 formal TPAS ratings 55.55%);

and high school (HS) teachers (51.82%) responded that the TPAS process was self-reflective for a teacher to either a “moderate amount” and/or to a “great deal.”

- The research did *not* confirm the initial hypothesis since NBCT did not report the highest percentage of participants that engaged in self-reflection; rather, teachers that were rated as “highly effective” actually possessed the highest percentage of respondents who claimed that they self-reflect (i.e., 81.20%). Ironically, it was the NBCT Cohort that reported the *lowest* percentage of teacher cohorts that engage in self-reflection (i.e., 67.00%). Middle school teachers were the only other teacher demographic besides “highly effective” teachers that reported to have more than 80% of their cohort engage in self-reflection (i.e., 80.30%)
- The results are much more in line with the researcher’s hypothesis regarding NBCT and their propensity to reflect as over 40% of all NBCT (i.e., 41.30%) reported that they feel that reflection impacts their teacher effectiveness “a great deal.”
- Overall, the largest percentage of responses of particular types of self-reflective activities consisted of teachers “*audio-taping*” their own teaching (16.297%); this was followed by teachers “*reviewing student performance data*” (15.678%); “*completing a self-reflective check-list*” (15.042%); and “*engaging in unstructured reflection*” (13.576%).
- There were a few particular self-reflective activities that presented a considerable delta between the NBCT and the non-NBCT cohort. “*Administering a student survey*” “*building a portfolio*”, and “*video-taping one’s class*” were three specific reflective activities that possessed the greatest margin of difference between these two cohorts.
- The data did not present large gaps in the *frequency* of self-reflective activities between the NBCT and the non-NBCT cohort. In fact, the NBCT and the non-NBCT cohort have

almost identical rates of reflective engagement, with over 90% of each cohort reporting that they engage in reflection - from a moderate to a large degree of frequency - for the two most popular self-reflective activities: “*engaging in unstructured reflection*” and “*reviewing student performance data.*”

- The data clearly demonstrated much higher degrees of self-reported frequency of the non-NBCT cohort in almost every different type of self-reflective activity; this was not at all expected.
- An overwhelming majority of teachers (85.37%) self-proclaim that they know how to “self-reflect” and do so often. Approximately 12% of survey respondents insist that they know how to “self-reflect” but choose not to since (they) “do not have the time.” Less than 1% of teachers self-report that they do not know how to self-reflect and/or they “know how to self-reflect” but do not believe that such actions assist instructional delivery.
- Middle school teachers were the only cohort that did not hit the 80% threshold of self-reported *self-reflection* (77.42%) and they also had the highest number of teachers’ claim that they did not know how to self-reflect.

The last research question – **Research Question Four (RQ4)** - scrutinized whether or not teachers that have been evaluated in the TPAS Framework grow in teacher effectiveness over time (as measured by their own TPAS rating) and what professional development factors most positively impact such growth. The researcher considered such things as the *total number of observations and most recent TPAS rating; years of service and TPAS rating; NBCT and TPAS rating; etc.* The data was unpacked using tabular crosstabs with regards to the *frequency of unstructured reflection and different demographical cohorts with most recent TPAS rating.* The

goal was to address which specific factors and viable professional development that most effectively builds teacher quality overtime.

The following abbreviated survey questions were utilized in answering Research Question 4 (RQ4):

- Assorted demographic data such as Q1(years of teaching); Q3(NBCT or not); Q5(highest degree conferred); Q6(placement level of teacher); Q12 (total number of TPAS observations in career) in relation to a teacher's first (Q13) and last (Q26) TPAS rating.
- Q24: Which of the following professional development activities has most positively impacted your instruction?
- Q25. As a teacher, how do you see yourself spending most of your "learning time" in the schoolhouse?

The results to RQ4 are as follows:

- There was not a positive relationship between the total number of observations and a teacher's last TPAS rating; thus undermining the claim that TPAS is a *formative* tool for building teacher effectiveness. If TPAS had been formative, then teachers that are subjected to more frequent formal TPAS observations, their average TPAS rated should summarily increase as well; this was clearly not the case.
- There is nothing definitive in the data that validates the long standing hypothesis (myth) that as years of teacher service increase; so, too, does one TPAS rating.
- NBCT have higher average TPAS ratings than their non-NBCT peers. The NBCT cohort has a significantly higher average TPAS rating (3.5833) as compared to the non-NBCT

cohort (3.4352) and the overall average TPAS rating (3.4455), respectively. The researcher cannot say that because they are NBCT they are superior teachers; only that there is a correlation, not a causation, with appreciable TPAS ratings.

- The researcher was surprised by the high number of teacher respondents overall who answered affirmatively that they - admittedly - self-reflect at least to a *moderate* degree (i.e., 80.7%), the data also speaks to the fact that one's commensurate average TPAS rating for these highly reflective practitioners is higher as well, albeit by only about 2% than their less than highly reflective practitioner peers.
- Similarly surprising to the researcher was that *degrees of reflection matter*. That is, the fact that teachers participate in reflection is not what impacts teacher effectiveness; rather, it seems it is the *degree/amount to which one reflects* that matters. For NBCT in particular, be a correlation between those teachers that *self-reflect to a large degree* and their corresponding average TPAS rating as evidenced by that cohort of teachers securing a 3.586 average TPAS rating, which was the highest recorded average TPAS rating amongst all demographic cohorts in this study.
- One out of every three NBCT do not reflect; but those NBCT that choose to, actually reflect to a large degree and feel strongly about its impact on their instruction.
- Overall, the non-NBCT cohort out-reflects the NBCT cohort.
- The more years of teaching is inversely related to the degree to which a teacher engages in self-reflection.
- Over 30% of teachers with more than 20 years of experience reported that they do not reflect at all on their teaching.

- Teachers with 16 to 20 years of experience had the highest TPAS average rating with respect to years of experience; NBCT had highest average TPAS rating of all teacher cohorts.
- Teaching experience seems to mute any lack of reflection in a teacher's TPAS rating.
- The degree to which elementary teachers reflect (i.e., large amounts of reported reflection) was the most; high school teachers reflected reported to have the lowest degree to which they self-reflect. Two out of every three middle school teachers reflect to at least a moderate degree.
- Unsure if the frequency of reflection in this data explains the TPAS rating for different cohorts.
- Less than 1% of various teacher respondent cohorts did not favorably respond to the "TPAS process and lesson observations" category as a viable professional development activity for the purposes of building teacher effectiveness as close to 1% of teachers in both the NBCT (0.90%) and the "highly effective" (1.0%) cohorts, similarly responded.
- For "highly effective" teachers, the highest percentage of respondents of most impactful professional development consisted of "*professional learning communities*" (21.90%); "*peer coaching and mentoring*" (20.0%); and, to a lesser degree, "*formal, credit-bearing coursework at a higher institution of learning*" (13.0%). "*Conferences*" (5.0%) and "*informal, continuing professional credit coursework*" (7.0%) were two of the lowest rated categories of professional development as reported by "highly effective" teachers (in conjunction with the aforementioned bullet, the "*TPAS process and lesson observation*" category was rated lowest by "highly effective" teachers).
- For the NBCT's cohort, the "*other*" category (21.90%) was rated by this group as most impactful in building their instructional prowess, followed by the act of "*engaging in self-reflection*"

(18.80%). What NBCT cited as “other” professional development in their write-in responses were as follows:

- “Informal shares with colleagues”
 - “National Board Certified process” (6 times)
 - “Time to plan individually and formal time set aside to team plan”
 - “Peer to peer walkthroughs”
 - “Online, self-guided professional development” (0.00%)
- For everyone else, there was a non-NBCT category (not including either NBCT or “highly effective” rated teachers) for the researcher to analyze. Similar to NBCT and the “highly effective” teachers cohort, non-NBCT responded that they were positively impacted by “*professional learning communities*” (21.90%) and “*peer coaching and mentoring*” (20.70%); these responses were consistent with the two aforementioned cohorts. What was surprising was that 21.0% of all non-NBCT felt that the “*TPAS process and lesson observations*” were impactful and helped to build their teacher effectiveness; this result was in direct contrast to the survey responses fielded by the NBCT and the “highly effective” teachers’ cohorts.

Recommendations to the District

The purpose of this study was to examine teachers’ perceptions of the impact of the local teacher evaluation system (TPAS) has on one’s effectiveness over time and to triangulate teacher evaluation, self-reflection, and their respective roles in improving teacher quality. The local school system in question can use the results to modify future observational practice /protocol and provide actionable data for helping to build local teacher prowess over time. Generally, speaking, central office and building level administrators can customize professional development topics based on teacher feedback in this study. Providing opportunities for topics that have been highlighted by teachers as most relevant and delivering the sessions in formats that teachers documented as most effective can increase teacher effectiveness through professional growth.

The results of the study will also help the local school system to prioritize principal and evaluator training and certification with a focus on professional growth. Certainly, the data from the study gives rise to the impetus of differentiated evaluation and support based on teachers' experience and past performance. Part of this support – locally – will be to facilitate the ability of evaluator and evaluatee to collaborate on areas of focused strengths and improvement for teachers. Support locally developed measures by engage teachers themselves in improving the local teacher evaluation system while pursuing improvements in the TPAS model creation and use, keeping in mind that the evaluation system should be able to “tell stories” that go beyond a teacher's performance rating.

Specifically, the researcher recommends the following to be implemented as a result of this research:

- Differentiated deployment of TPAS (formative vs. summative)
- A *formative* evaluation of teachers should encompass more explicit reflection similar to NBCT teachers experiences (such as video-tape review and/or portfolio build) to also include peer review but conducted without a formal TPAS observation
- Encourage more teachers to pursue National Board Certification (NBC)
- More time and resources (human capital) need to be devoted to improving the TPAS system.
- A slight modification of TPAS overall process with a focus on the following:
 - Evaluators and evaluatees should more forcibly unpack TPAS' *Domain 3: Instruction* (and, to a lesser extent, *Domain 2: Classroom Management*)
 - Devote more attention and resources to the *post-conference*

- Allocate more informal observation in the form of *walk-throughs* and *peer observations*
- Use more *virtual, turn-key professional development* (such as Performance Matters' Professional Development suite) to build teaching capacity
- To ensure more inter-rater reliability, the following steps should be taken by the local school system:
 - Create a cyclical pool of full time "TPAS" administrators from current principals and supervisors) at central office to work on TPAS more formal training for new/experienced administrators to build their capacity as evaluators (three year commitment to the position).
 - These TPAS administrators would liaison with school based administrators to conduct formal and informal TPAS observations in the field; this would address some of the administrator change at the building level and inconsistency with how the evaluation tool is deployed across the local school system.
 - These TPAS administrators would embed more explicit self-reflective activities that present "highly effective" and NBCT most enthusiastically endorse to build teacher effectiveness over time; specifically, reviewing video tape of one's teaching; engaging in unstructured reflection; peer observation; reviewing students' performance data on a daily/weekly basis; and building a dynamic teacher portfolio.
 - Conduct an annual review of the TPAS Observation Rubric with assorted stakeholders across the school system

The overall study validates most of what this researcher had hypothesized with regards to the benefit of self-reflection and how teachers – as reflective practitioners – can build their own teacher effectiveness over time by engaging in such behavior. At the same time, the researcher was struck by the amount of reflection in which all teachers engage agnostic of demographic. Most specifically, the “highly effective” teacher cohort responded with the highest level of self-reflection – even higher than NBCTs’ had reported. Hence, how can the system continue to engage more teachers to “self-reflect” – both in structured and in unstructured ways? At the same time, the research gives rise to minor modifications to how TPAS could be deployed across the school system so as to make the teacher evaluation process truly formative (see above).

Suggestions for Future Research:

An initial suggestion for future research would reside in any investigation to include additional t-tests to similar queries for all disaggregated teacher cohorts. Additionally, researchers should further confirm these findings with more analytical statistical significance. Furthermore, a complement to this present research study would be to examine whether or not tenured teachers who are evaluated with TPAS actually improve their teacher effectiveness over time (as measured by TPAS). This recommended research would be accomplished via a records review of all SMCPS’ TPAS ratings for tenured teacher cohorts – also differentiated by various demographic data in the school district - over a three year period; this data would emanate from actual TPAS ratings (and results would not rely on self-reporting from teachers). From this analyses, researchers could identify if general data themes emerge for varying experiences, background, education, and/or training (such as National Board Certification (NBC) teachers, non-NBC teachers, teachers with more than 20 years of experience, etc.). In other words, are there differences in the "travel times “for assorted teacher cohorts – per one’s own particular

demographic – for “effective” teachers moving to "highly effective" within the TPAS framework?

It should also be noted that the “act of self-reflection” and/or “engaging in self-reflective activity” could mean something slightly different to different people/cohort. So when teacher cohorts had responded to the survey, there may have been varied interpretations as to what self-reflection actually looks like and what activities personify a “self-reflection.” Moreover, some individuals may actually “reflect” without even realizing that they are reflecting given that the act is deeply embedded in one’s repertoire. For future research, self-reflective acts should be defined in terms of the following four modes of reflective thinking that Lana Danielson (2009) defined and previously cited in this research:

1. *Technological (or Formulaic) Thinking*
2. *Situational Thinking*
3. *Deliberate Thinking*
4. *Dialectical Thinking*

Specifically, a second, follow-up phase to this study would be to use three years of teacher evaluation data (for the following school years (SY) SY2013-14, SY2014-5 and SY2015-6, respectively) from the school system, which could be collected as part of the current teacher evaluation system (TPAS) to measure the growth of a teacher over time per their TPAS observation/evaluation rating.. Moreover, the research could consider teacher evaluation/performance in the form of TPAS ratings derived from several measures (such as final indicator level observation scores, average overall observation scores, and summative evaluation scores) and use system demographic data to see if general data themes emerge for various teacher cohorts such as NBC teachers, non-NBC teachers, teachers with more than 20

years of experience, etc. The major emphasis in this records reflection is to determine whether or not NBC teachers – in general – are more highly rated than non-NBC teachers are on TPAS. Additionally, this researcher will identify the average “travel times” of both NBC teachers and non-NBCT as teacher ratings move from “effective” to “highly effective” so as to determine if there are in fact significant differences in the travel times for each aforementioned teacher cohort. Notwithstanding, by including only tenured teachers, this framework would be consistent with the data collection that occurred in the survey and all future research should work to mitigate the effect of a non-tenured teacher in the records reflection by eliminating all non-tenured teachers in the review since only tenured teachers are eligible to be considered for NBCT. Lastly, using a records reflection in conjunction with the survey results would help to rely upon only the “self-reporting” of one’s TPAS.

APPENDICES

Glossary

Key terms, seminal to this study encompass the following:

National Board for Professional Teaching Standards (NBPTS): A board made up of a regional and state membership structure that establishes standards for what master teachers should know and be able to do through student results regarding academic achievement, career readiness, and personal responsibility.

National Board Certified Teacher (NBCT): A teacher who meets high and rigorous standards established by the NBPTS through rigorous independent study, self-reflection and assessment, and evaluation by the NBPTS.

Non-National Board Certified Teacher (non-NBCT): A teacher who has not met the high and rigorous standards established by the NBPTS.

National Board Certification (NBC): a certificate issued to signify a teacher has met the requirements of the NBPTS and is a NBCT.

Common Core of State Standards (CCSS): a research and evidence-based, internationally benchmarked, set of shared educational standards of goals and expectations of what students should understand and be able to do in grades K-12 in order to be successfully career and college-ready.

Teacher: defined as any teacher holding full tenure in the local school system

Teacher self-efficacy: a teacher's confidence in their ability to increase students' academic achievement, career readiness, and personal responsibility through effective instruction and evaluation.

Teacher Sense of Efficacy Scale (TSES): a widely used Likert-like instrument developed by Tschannen-Moren and Woolfolk and Hoy that measures a teacher's sense of teaching self-efficacy, providing a means to capture and measure the perceived efficacy in the areas of student engagement, classroom management, and instructional strategies.

Self-assessment: Airasian and Gullickson (1994) define self-assessment as "the process of making judgments about the appropriateness or effectiveness of one's own knowledge, performance, beliefs, products, or effects, so that they can be improved or refined" (p. 6). Importantly, the teacher's knowledge utilized during self-assessment must be articulated to determine consciously what needs to change and consider the steps that will be taken to refine future teaching skills or understanding.

Reflection: Reflection, in this study, refers to the thinking teachers do about their practice—this thinking could be done during teaching (reflection-in-action) or following the teaching episode (reflection-on-action) (Schôn, 1983). Reflection also refers to the metacognitive processes teachers employ to compare, evaluate, and provide direction for their teaching practices (Calderhead, 1989; Ertmer & Newby, 1996). Researchers often describe the tacit nature of teachers' reflections (Grimmett, 1988; Richardson, 1990) and the use of feelings, intuition, or trial and error to make educational decisions (Grimmett & MacKinnon, 1992). This study does not dispute that teachers' reflections can be tacit, but for reflection to inform and change teaching practices most effectively, it is necessary for teachers to compare their classroom practices explicitly to those practices advocated by the education community. Thus, in this study reflection refers to teachers' explicit thoughts concerning teaching.

Appendix A

District Framework for Teachers: Rubrics for Domains 2 and 3 included only

The Framework for the Teacher Performance Assessment System

The narratives and rubrics in this section have been adapted from Charlotte Danielson's *Enhancing Professional Practice: A Framework for Teaching, 2nd Ed.* (2007, ASCD). The domains and components align with the *Maryland Teacher Evaluation Framework*.

Domain 1	Planning and Preparation
Component 1a	Demonstrates Knowledge of Content and Pedagogy
Component 1b	Demonstrates Knowledge of Students
Component 1c	Selects Instructional Outcomes
Component 1d	Demonstrates Utilization of Resources
Component 1e	Designs Coherent Instruction
Component 1f	Assesses Student Learning
Domain 2	The Learning Environment
Component 2a	Establishes an Environment of Respect and Rapport
Component 2b	Establishes a Culture for Learning
Component 2c	Manages Classroom Procedures
Component 2d	Manages Student Behavior
Component 2e	Organizes Physical Space
Domain 3	Instruction
Component 3a	Communicates Clearly and Accurately
Component 3b	Uses Higher Order Questioning and Discussion Techniques
Component 3c	Engages Students in Learning
Component 3d	Uses Assessment in Instruction
Component 3e	Demonstrates Flexibility and Responsiveness
Domain 4	Professional Responsibilities
Component 4a	Grows and Develops Professionally
Component 4b	Communicates with Families
Component 4c	Participates in a Professional Learning Community
Component 4d	Shows Professionalism
Component 4e	Maintains Accurate Records
Domain 5	Evidence of Student Learning
Component 5a	Summative Assessments
Component 5b	Formative Assessments
Component 5c	Performance Assessments
Component 5d	Growth Assessments

Appendix B
The Teacher Performance Assessment System (TPAS) Summary for Evaluation Cycle
Requirements Sample Materials
Timelines and Guidelines for Submission

The following chart summarizes evaluation cycle requirements. Documents, artifacts and reflections are uploaded via the online TPAS module, <http://eval.smcps.org>

Summative Non-Tenured Meeting Requirements

Activities	Teacher Completes	Evaluator Completes	Wheretostore/send?
Initial conference establishing SLO's <i>by October 15</i>	Self-reflection; SLO Worksheet		Submitted via the online TPAS module under "Student Learning"
Observations - (4 observations during the year*). Each observation includes a pre-conference & a post-conference <i>First observation cycle by winter break; Second formal observation cycle by April 15</i>	Lesson plan to each pre-observation conference. Reflection following the observation.	Observation write-up; discussed at post-conference	Submitted via the online TPAS module under "Observations"
Mid-year evaluation and Review of SLOs <i>by January 30</i>	Artifacts, documentation as needed (e.g., parent contact logs). SLO Summary/Reflection	Evaluation; Domain 5 – Student Learning Record	Evaluation – TPAS under "Evaluation"; signed evaluation is sent to Human Resources; Domain 5 Performance Assessment under "Student Learning"
End-of-Year Evaluation and Review of SLOs <i>by the last teacher workday.</i>	Artifacts, documentation as needed (e.g., parent contact logs). SLO Summary/Reflection	Evaluation; Domain 5 – Student Learning Record	Evaluation – TPAS under "Evaluation"; signed evaluation is sent to Human Resources; Domain 5 Performance Assessment under "Student Learning"

*Evaluators will make every effort possible to schedule observations with the appropriate supervisor to condense the four scheduled observations to two; if a school based administrator and supervisor conduct an observation together, it counts as two of the required four.

Summative Tenured Meeting Requirements

Activities	Teacher Completes	Evaluator Completes	Wheretostore?
Initial conference establishing SLO's <i>by October 15</i>	Self-reflection; SLO Worksheet		TPAS under "Student Learning"
2 observations during the year; each observation includes a pre-conference & a post-conference <i>by May 15</i>	Lesson plan to each pre-observation conference. Reflection following the observation.	Observation write-up; discussed at post-conference	TPAS under "Observations"
Mid-year review of SLO data <i>by January 30</i>	SLO Summary/Reflection		Domain 5 Performance Assessment under "Student Learning"
Evaluation by the end of the year <i>by the last teacher workday.</i>	SLO Summary/Reflection	Evaluation; Domain 5 – Student Learning Record	Evaluation – TPAS under "Evaluation"; signed evaluation is sent to Human Resources; Domain 5 Performance Assessment under "Student Learning"

Formative Meeting Requirements for the 2015-2016 School Year:

Activities	Teacher Completes	Evaluator Completes	Wheretosend/store?
Initial conference establishing SLO's <i>by October 15</i>	Self-reflection; SLO Worksheet		TPAS under "Student Learning"
Mid-year review of SLO data <i>by January 30</i>	Domain 5 Performance Assessment		Domain 5 Performance Assessment under "Student Learning"
Evaluation by the end of the year <i>by the last teacher workday.</i>	TPAS Formative Assessment Summary; SLO Summary/Reflection	Domain 5 – Student Learning Record	TPAS Formative Assessment Summary under "Artifact"; a signed copy is sent to Human Resources; Domain Five Performance Assessment under "Student Learning"

Appendix D – Implied Informed Consent Informed Consent Form

Purpose

This study attempts to quantify teacher perception of the teacher evaluation metric (Teacher Performance Assessment System - TPAS) that is used to rate teachers. The research is being conducted by Alex Jaffurs under the supervision of Dr. Patricia Richardson at the University of Maryland, College Park. We are inviting you to participate in this research project because you are a tenured teacher in St. Mary's County.

Procedures

Your participation in the Teacher Perception Survey would include a 10-minute web-based questionnaire. The survey asks questions about the teacher evaluation process, the teacher evaluation metric, teacher routines/repertoire, and demographics. For the teacher perception questions you will be asked to comment on your own professional opinion of the benefit(s) that being evaluated in TPAS system – drilled down to each instructional Domain - bestow upon you as a teacher. Additionally, you will be asked about the teacher evaluation process and other factors that contribute to improved teacher performance over time There will also be a records review that will precede this survey for any tenured teacher in the county over the last three years (which would include the 2013-4, 2014-5; and 2015-6 School Years, respectively).

Risks/Discomforts

There are no more than minimal risks known to participants. In order to prevent breach of confidentiality, your responses will be coded and anonymous.

Benefits

There are no direct benefits for individual participants. However, it is hoped that through your participation, researchers will learn more about the flow of communication in elementary schools.

Confidentiality

All data obtained from participants will be kept confidential and will only be reported in an aggregate format (by reporting only combined results and never reporting individual ones). All questionnaires will be concealed, and no one other than the primary investigator listed below will have access to them. The data collected will be stored in the HIPPA-compliant, secure database until it has been deleted by the primary investigator.

Incentive

There will be no incentive to completing the survey.

Informed Consent Page 1 of 2

Appendix D – Implied Informed Consent (Continued)

Participation

Participation in this research study is completely voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your employment status in the county. If you desire to withdraw, please close your Internet browser.

Questions about the Research

If you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the investigator:
Alex Jaffurs, at 301-481-4639 or ajaffurs@umd.edu.

Questions about your Rights as Research Participants

If you have questions you do not feel comfortable asking the researcher, you may contact (Dr. Patricia Richardson, Mentoring Professor), 3119 Benjamin Building, pmr20659@gmail.com.

I have read, understood, and printed a copy of, the above consent form and desire of my own free will to participate in this study.

Yes

No

Appendix E

Teacher Survey Letter

Dear Teachers,

Thank you for taking the time to complete this survey. Your participation in the [Teacher Survey](#) will include a 10-minute web-based questionnaire about the TPAS process. The survey will also record identifiable data regarding your professional background (level of education, experiences, etc.).

You were selected for this [survey](#) because you have been formally rated at least once in TPAS.

This survey is anonymous and identifying information will not be collected or stored. General demographic information will be collected but will not be used for identification purposes, only to correlate similarities and differences in survey responses. After completing the demographic information, questions will focus on the following response areas:

- Perceptions of TPAS and its impact on the quality of your teaching.
- Perceptions of engaging in self-reflective behaviors and its impact on the quality of your teaching.

As you respond to questions, please reflect on your personal experience as it relates to the TPAS framework and your propensity for engaging in self-reflection. This study attempts to gauge the ways in which the school system's teachers believe that TPAS (Teacher Performance Assessment System) and/or the engagement of self-reflective behaviors impact the quality of their teaching.

The link to the survey can be found [HERE](#).

I appreciate your time and responses in completing this survey.

Regards,
Alex Jaffurs

Appendix F

Qualtrics Teacher Survey for Research Questions

Survey for Research Questions

1. How long have you been employed as an educator in your career?

- Less than one year (1)
- 1 to 2 years (2)
- 3 to 5 years (3)
- 6 to 10 years (4)
- 11 to 15 years (5)
- 16 to 20 years (6)
- more than 20 years (7)

2. Are you presently tenured in the St. Mary's County Public Schools (SMCPS) as a teacher?

- Yes (1)
- No (2)

3. Do you presently hold credentials as a National Board Certified Teacher (NBCT) in your field of expertise?

- Yes (1)
- No (2)

4. Are you currently in the process of pursuing National Board Certification (NBC) in your field of expertise?

- Yes (1)
- No (2)

5. What is the highest degree that you have been conferred?

- Undergraduate degree in teaching (e.g., elementary education) (1)
- Undergraduate degree in field of expertise (e.g., mathematics) (2)
- Master's degree (not in education) (3)
- Master's degree (education) (4)
- Doctoral degree (5)

6. Which of the following would best describe you as a teacher?

- Elementary School Teacher only (1)
- Middle School Teacher only (2)
- High School Teacher only (3)
- Elementary and Middle School Teacher (K to 8th grade) (4)
- Secondary School Teacher (Middle and High School teacher) (5)

7. How long have you taught at your current school?

- One year or less (1)
- Two to five years (2)
- Six to nine years (3)
- Greater than 10 years (4)

8. Have you changed schools at all over the last three consecutive school-years?

- Yes (1)
- No (2)
- Not applicable (4)

9. Has your local administrative team (principal and/or assistant principal) changed over the last three years?

- Yes (1)
- No (2)
- Not applicable (3)

10. Has your content supervisor changed over the last three years?

- Yes (1)
- No (2)
- Not applicable (3)

11. Approximately, how many times in the current school year have you been formally observed using the Teacher Performance Assessment System (TPAS) in SMCPS by at least one administrator?

- Zero (1)
- One time (2)
- Two times (3)
- Three times (4)
- Four or more times (5)

12. Approximately, how many times in your entire career in SMCPS have you been formally observed using the Teacher Performance Assessment System (TPAS) by at least one administrator?

- Never (1)
- 1 to 2 times (2)
- 3 to 5 times (3)
- 6 to 10 times (4)
- 11 to 15 times (5)
- 16 to 20 times (6)
- More than 20 times (7)

13. To the best of your knowledge, what was your overall TPAS rating on the very first formal observation you ever received from an administrator in SMCPS?

- Ineffective (1)
- Developing (2)
- Effective (3)
- Highly Effective (4)
- Do not remember (5)

14. In your professional opinion, which singular TPAS Domain most directly impacts your teaching skill?

- Domain 1: Planning and Preparation (1)
- Domain 2: The Learning Environment (2)
- Domain 3: Instruction (3)
- Domain 4: Professional Responsibilities (4)

15. In your professional opinion, which singular TPAS Domain is least likely to impact your teaching skill?

- Domain 1: Planning and Preparation (1)
- Domain 2: The Learning Environment (2)
- Domain 3: Instruction (3)
- Domain 4: Professional Responsibilities (4)

16. In your professional opinion, can you rank order below each TPAS Domain that impacts your teaching skill from least effective to most effective?

- _____ Domain 1: Planning and Preparation (1)
- _____ Domain 2: The Learning Environment (2)
- _____ Domain 3: Instruction (3)
- _____ Domain 4: Professional Responsibilities (4)

17. As an educator, what statement below best describes your professional opinion about the TPAS process and using TPAS as a tool for teacher evaluation?

- TPAS is a tool that is used by SMCPS administrators for compliance only (1)
- TPAS is a tool that is used by SMCPS administrators for teacher quality control only (2)
- TPAS is a tool that is used by SMCPS administrators for both compliance and teacher quality control (6)
- TPAS is a tool that that is used by SMCPS teachers as a professional development tool to improve their classroom instructional delivery only (3)
- TPAS is a tool that is used by SMCPS administrators for compliance, teacher quality control, and professional development (4)
- None of the above (7)

18. In your professional opinion, what has been the impact of the TPAS process on your development as a classroom teacher?

- No impact (1)
- A small impact (2)
- A moderate impact (3)
- A large impact (4)

19. In your professional opinion, which part of the TPAS process has positively impacted your teaching the most as a teacher in SMCPS?

- The pre-conference (1)
- The lesson planning (2)
- The actual observation (3)
- The post-conference (4)
- Other (5) _____

20. In your professional opinion, how much of the TPAS process engages a teacher in behaviors that are self-reflective?

- None at all (1)
- A little (2)
- A moderate amount (3)
- A great deal (4)

21. In your professional opinion, which of the following descriptors best describe your experience of the TPAS process up to this point as a classroom teacher in SMCPS?

- Lack of engagement (1)
- Poor use of time (2)
- Poorly organized/executed (3)
- Reflective in nature (4)
- Energizing (5)
- Supportive (6)
- Impactful (7)

22. In your professional opinion, which of the following descriptors best describe the ideal state of the TPAS process as it could be as a classroom teacher in SMCPS?

- Lack of engagement (1)
- Poor use of time (2)
- Poorly organized/executed (3)
- Reflective in nature (4)
- Energizing (5)
- Supportive (6)
- Impactful (7)

23. How would you rate the TPAS process in your development as a teacher with respect to the following professional duties?

	Very satisfied (1)	Somewhat satisfied (2)	Neutral (3)	Somewhat dissatisfied (4)	Very dissatisfied (5)
Dealing with the day to day "workflow" (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Implementing the Common Core(MCCRS) (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using student data (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Engaging in self-reflection and self-assessment (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Integrating educational technology (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Differentiating instruction (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lesson planning and using curricular resources (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Building overall teaching capacity (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Working in a professional learning community (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Classroom management (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

24. Which of the following professional development activities has positively impacted your instruction the most?

- Formal, credit-bearing coursework at a higher institution of learning (1)
- Informal, continuing professional credit coursework (2)
- The TPAS process and lesson observations (3)
- Peer coaching and mentoring (4)
- Engaging in self-reflective activities (12)
- Professional learning communities (5)
- System-wide professional development (6)
- Online, self-guided professional development (7)
- Conferences (9)
- Intensive summer trainings (10)
- Other (11) _____

25. As a teacher, how do you see yourself spending most of your "learning time" in the schoolhouse?

- Meetings (1)
- Observations and evaluations (2)
- Professional development (3)
- Testing (4)
- Forms (5)
- Preparing to teach (8)
- Actual teaching (7)
- Reflecting on teaching (6)
- Other (9) _____

26. Generally speaking, which of the following most accurately reflects your most recent TPAS rating you have received?

- Ineffective (1)
- Developing (2)
- Effective (3)
- Highly Effective (4)

27. For each question below, please mark one choice in part (A). If you answer "Yes" in part (A) then please mark one choice in part (B) to indicate how much impact it had upon your development as a teacher. Did you participate in any of the following self-reflective behaviors, and what was the impact of these self-reflective behaviors on your development as a teacher?

	(A) Participation		(B) Impact			
	Yes (1)	No (2)	No impact (1)	A small impact (2)	A moderate impact (3)	A large impact (4)
Audio-taping one's teaching (1)	<input type="radio"/>					
Video-taping one's teaching (2)	<input type="radio"/>					
Journal writing (3)	<input type="radio"/>					
Building a portfolio (4)	<input type="radio"/>					
Completing a self-reflection checklist (daily/weekly) (5)	<input type="radio"/>					
Reviewing Student Performance Data (daily/weekly) (6)	<input type="radio"/>					
Administering a student survey(s) to your class (7)	<input type="radio"/>					
Engaging in unstructured reflection (e.g., brainstorming) (8)	<input type="radio"/>					

28. For each question below, please mark one choice in part (A). If you answer "Yes" in part (A) then please mark one choice in part (B) to indicate how often as a teacher you had utilized such reflection. Did you participate in any of the following self-reflective behaviors, and how often did you personally engage in these self-reflective behaviors?

	(A) Participation		(B) Frequency		
	Yes (1)	No (2)	A little (1)	A moderate amount (2)	A great deal (3)
Audio-taping one's teaching (1)	<input type="radio"/>				
Video-taping one's teaching (2)	<input type="radio"/>				
Journal writing (3)	<input type="radio"/>				
Building a portfolio (4)	<input type="radio"/>				
Completing a self-reflection checklist (daily/weekly) (5)	<input type="radio"/>				
Reviewing Student Performance Data (daily/weekly) (6)	<input type="radio"/>				
Administering a student survey(s) to your class (7)	<input type="radio"/>				
Engaging in unstructured reflection (e.g., brainstorming) (8)	<input type="radio"/>				

29. As a teacher, I generally know how to "self-reflect" on my own practice.

- Yes, I do know how to "self-reflect" and do so often. (1)
- Yes, I do know how to "self-reflect" but choose not to because I do not feel that it helps my teaching. (2)
- Yes, I do know how to "self-reflect" but choose not to since I do not have the time. (3)
- Yes, I do know how to "self-reflect" but choose not to for other reasons. (4)
- No, I do not know how to self-reflect. (5)

References

- Achinstein, B., & Ogawa, R. (2006). (in)fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, 76(1), 30-63. Retrieved from <http://www.metapress.com/content/E14543458R811864>
- Andrejko, L. (Fall 1998). The Case for the Teacher Portfolio. *Journal of Staff Development* 19(4).
- Attinello, J. R., Lare, D. and Waters, F. (June 2006). The Value of Teacher Portfolios for Evaluation and Professional Growth. *NASSP Bulletin* 90(2).
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, 50(2), 248-287.
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-efficacy: The exercise of control. *Journal of Cognitive Psychotherapy*, 13(2), 158-166.
- Barnes, G. L. V. (1998). *A comparison of self-efficacy and teaching effectiveness in preservice string teachers* (Doctoral dissertation, The Ohio State University).
- Bellafante, G. (2012, January 12). Petty Differences Mask Consensus on Teachers. The New York Times. Retrieved April, 2017, from http://www.nytimes.com/2012/01/22/nyregion/more-agreement-than-disagreement-on-how-to-assess-teachers.html?_r=1
- Berg, J. H. (2005, April 15). *Board certification during teaching's second stage: Professionalizing teaching through differentiated roles*. Paper presented at the American Educational Research Association Annual Conference, Montreal.
- Best Teachers Have Sense of Self-Efficacy. (n.d.). Retrieved April 15, 2016, from <http://www.ernweb.com/educational-research-articles/teacher-self-efficacy-effectiveness/>
- Bill & Melinda Gates Foundation (2010). Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project (Research Paper). Seattle, WA: Author. Retrieved December 16, 2010, from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf.
- Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M. and Hess, J. (2007). Examining District Guidance to Schools on Teacher Evaluation Policies in the Midwest Region. *Learning Points Associates-REL Midwest* Retrieved on January 4, 2011, from http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2007030.pdf

- Brandt, C., Thomas, J. and Burke, M. (2008). State policies on teacher evaluation practices in the Midwest Region. *Midwest Regional Educational Laboratory* Retrieved on November 11, 2010, from http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/techbrief/tr_00408.pdf
- Braun, H. I. (2005). Using Student Progress To Evaluate Teachers: A Primer on Value-Added Models. *Educational Testing Service* Retrieved on March 23, 2011, from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>
- Bruce, M. (2010) Controversy Surrounds White House Push For Increased Teacher Accountability. *ABC News July 29, 2010* Retrieved on February 1, 2011, from <http://abcnews.go.com/Politics/controversy-surrounds-white-house-push-increased-teacher-accountability/story?id=11279505&page=1>
- Camera, L (2014) Waiver States Can Seek a Year’s Delay in Using Tests for Teacher Evaluation,” *Education Week*, August 21, 2014, available from http://blogs.edweek.org/edweek/campaign-k-12/2014/08/ed_announcement.html
- Cavalluzzo, L., Barrow, L., Henderson, S., Mokher, C., Geraghty, T., & Sartain, L. (2014). From large urban to small rural schools: An empirical study of National Board Certification and teaching effectiveness. Arlington, VA: CNA Education.
- Chambers, J., Brodziak de los Reyes, I., and O’Neil, C., *How Much Are Districts Spending to Implement Teacher Evaluation Systems? Case Studies of Hillsborough County Public Schools, Memphis City Schools, and Pittsburgh Public Schools*. American Institutes for Research: RAND Education and American Institutes for Research, May 2013
- Cogshall, J. G., Rasmussen, C., Colton, A., Milton, J., & Jacques, C. (2012). Generating Teaching Effectiveness: The Role of Job-Embedded Professional Learning in Teacher Evaluation. Research & Policy Brief. National Comprehensive Center for Teacher Quality.
- Cowan, J., & Goldhaber, D. (2015). National Board certification and teacher effectiveness: Evidence from Washington. *The Center for Data & Research, University of Washington Bothell*.
- City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education: A network approach to improving teaching and learning*. Cambridge, MA: Harvard University Press.
- Clift, R. T., Houston, W. R., & Pugach, M. C. (Eds.). (1990). Encouraging reflective practice in education: An analysis of issues and programs. New York: Teachers College Press.
- Constantino, P. M., & De Lorenzo, M. N. (2001). Developing a professional teaching portfolio: A guide for success. Boston: Allyn and Bacon.

- Criteria for an Effective Teacher Evaluation System. (2014). Retrieved June 15, 2016, from <http://www.aft.org/periodical/american-educator/spring-2014/criteria-effective-teacher-evaluation-system>
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. ASCD.
- Danielson. (2009). Fostering Reflection. *Educational Leadership*, 66(5). Retrieved February 21, 2017, from <http://www.ascd.org/publications/educational-leadership/feb09/vol66/num05/Fostering-Reflection.aspx>
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives* 8(1). Available: <http://olam.ed.asu.edu/epaa/v8n1/>
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- Darling-Hammond, L., Jaquith, A., & Hamilton, M. (2012). Creating a comprehensive system for evaluating and supporting effective teaching. *Stanford, California: Stanford Center for Opportunity Policy in Education (SCOPE)*. Retrieved October, 26, 2012.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. New York: D.C. Heath.
- Dewey, J. (1938). *Experience and education*. New York: MacMillan.
- Dougherty, K. M. & Jacobs, S. (2015, November). State of the states 2015: Evaluating, teaching, leading, and learning. Washington, D.C.: National Council on Teacher Equality, as cited in Pennington, K. & Mead, S. (2016, December). For good measure?: Teacher evaluation policy in the ESSA era. Washington, D.C., Bellwether Education Partners. Retrieved January 2, 2017, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- Doyle, D. & Han, J. G. (2012). *Measuring teacher effectiveness: A look “under the hood” of teacher evaluation in 10 sites*. (New York: 50CAN; New Haven, CT: ConnCAN; and Chapel Hill, NC: Public Impact, 2012), available at <http://www.conncan.org/learn/research/teachers/measuring-teacher-effectiveness>
- DQC. (2011). *DQC 2011 Annual Survey Data Quality Campaign* Retrieved on February 16, 2011, from <http://www.dataqualitycampaign.org/survey>

- Duke, D.L., & Stiggins, R.J. (1986). *Teacher evaluation-Five keys to growth*. Washington, DC: National Education Association.
- ESEA Flexibility. (2016, May 12). Retrieved April, 2017, from <https://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>
- Ellett, C., & Teddlie, C. Teacher Evaluation, Teacher Effectiveness and School Effectiveness: Perspectives from the USA. [Journal of Personnel Evaluation in Education](#) March 2003, Volume 17, [Issue 1](#), pp 101-128
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.
- EPI. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. *Economic Policy Institute* Retrieved on November 1, 2010, from http://epi.3cdn.net/b9667271ee6c154195_t9m6ijj8k.pdf
- Fiarman, S. E. (2007, April). *It's hard to go back: Career decisions of second-stage teacher leaders*. Paper presented at the Annual Conference of the American Educational Research Association, Chicago, IL.
- Fiarman, S. E., Johnson, S. M., Munger, M. S., Papay, J. P., & Qazilbash, E. K. (2009). *Teachers leading teachers: The experiences of peer assistance and review consulting teachers*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.
- Frank, K. A., Sykes, G., Anagnostopoulos, D., Cannata, M., Chard, L., Krause, A., & McCrory, R. (2008). Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? *Educational Evaluation and Policy Analysis*, *30*(1), 3-30. doi: 10.3102/0162373707313781.
- Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education*, 0022487114542519.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915-945.
- Gill, B., English, B., Furgeson, J., McCullough, M. Institute of Educational Sciences retrieved 6/20/2014 http://mathematica-mpr.com/publications/pdfs/education/REL_StudentGrowthMeasures.pdf
- Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raundenbush, S. and Whitehurst, G. (2010). Evaluating Teachers: The Important Role of Value-Added. *Brown Center on Education Policy at Brookings* Retrieved on January 5, 2011, from http://www.brookings.edu/reports/2010/1117_evaluating_teachers.aspx

- Glickman, C. D. (2002). *Leadership for learning: How to help teachers succeed*. Alexandria, VA: ASCD.
- Goldhaber, D. (2010). When the Stakes Are High, Can We Rely on Value-Added? *Center for American Progress* Retrieved on January 21, 2011, from <http://www.americanprogress.org/issues/2010/12/vam.html>
- Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, 89(1), 134-150.
- Goldhaber, D. and Hansen, M. (2008). Is it Just a Bad Class? Assessing the stability of measured teacher performance. *Center for Reinventing Public Education* Retrieved on November 5, 2010, from http://www.crpe.org/cs/crpe/download/csr_files/wp_crpe5_badclass_nov08.pdf
- Goldstein, J., & Noguera, P. A. (2006). A thoughtful approach to teacher evaluation. *Educational Leadership*, 63(6), 31.
- Grossman, P. (2008). Responding to our critics: From crisis to opportunity in research on teacher education. *Journal of Teacher Education*, 59(1), 10-23. doi: 10.1177/0022487107310748.
- Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D. and Lankford, H. (2010). Measure for Measure: The relationship between measures of instructional practice in middle school English Language Arts and teachers' value-added scores. *CALDER* Retrieved on November 2, 2010, from http://www.caldercenter.org/upload/CALDERWorkPaper_45.pdf
- Grossman, P. L., Wineburg, S., & Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College Record*, 103(6), 942-1012.
- Hargreaves, A., & Fink, D. (2009). Distributed leadership: Democracy or delivery? In A. Harris (Ed.), *Distributed leadership: Different perspectives* (Vol. 7, pp. 181-193). London: Springer.
- Grimmett, P. P., Erickson, G. L., Mackinnon, A. A., & Riecken, T. J. (1990). Reflective practice in teacher education. In R. T. Clift, W. R. Houston, & M. C. Pugach (Eds.), *Encouraging reflective practice in education: An analysis of issues and programs* (pp. 20–38). New York: Teachers College Press.
- Hai-Jew, S. (2017, December). Conducting a Cross Tabulation Analysis in the Qualtrics Research Suite. Retrieved October 12, 2017, from <http://scalar.usc.edu/works/c2c-digital-magazine-fall-2016--winter-2017/conducting-a-cross-tabulation-analysis-qualtrics-research-suite>
- Hanushek, E. A. (2010). The Economic Value of Higher Teacher Quality. *National Bureau of Economic Research* Retrieved on January 25, 2011, from <http://papers.nber.org/papers/w16606>
- Hargreaves, A., & Fullan, M. G. (1992). *Understanding teacher development*. New York: Teachers College Press.

- Harris, D. N. and Sass, T. R. (2009). What Makes for a Good Teacher and Who Can Tell? *CALDER* Retrieved on February 3, 2011, from <http://www.urban.org/UploadedPDF/1001431-what-makes-for-a-good-teacher.pdf>
- Hatch, T., White, M. E., & Faigenabum, D. (2005). Expertise, credibility, and influence: How teachers can influence policy, advance research, and improve performance. *Teachers College Record*, 107(5), 1004-1035.
- Hill, H. C. (2015). Review of the Mirage: Confronting the Hard Truth about our Quest for Teacher Development.
- Hull, J. (2011). Building a better evaluation system: Full report. Retrieved March 03, 2015, from <http://www.centerforpubliceducation.org/Main-Menu/Staffingstudents/Building-A-Better-Evaluation-System/Building-A-Better-Evaluation-System.html>
- Hull, J. (2013, March). *Trends in teacher evaluation: How states are measuring teacher performance*. Retrieved from <http://www.centerforpubliceducation.org/Main-Menu/Evaluatingperformance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-EvaluationFull-Report-PDF.pdf>
- Hunzicker, J. (2012). Professional development and job-embedded collaboration: How teachers learn to exercise leadership. *Professional Development in Education*, 38(2), 267-289.
- Ikemoto, G., Taliaferro, L., & Adams, E. (2012). Playmakers: how great principals build and lead great teams of teachers. *New York: New Leaders Institute*.
- Jacob, A., & McGovern, K. (2015). The Mirage: Confronting the Hard Truth about Our Quest for Teacher Development. *TNTP*.
- Jacob, B. A. and Lefgren, L. (January 2008). Can Principals Identify Effective Teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics* 26(1).
- Jacobs, S., Brody, S., Rosen, K. M. and Madden, T. M. (2011). 2010 State Teacher Policy Yearbook. *National Council on Teacher Quality* Retrieved on January 31, 2011, from <http://www.nctq.org/stpy09/updates/primaryFindings.jsp>
- Jerald, C. D. and Hook, K. V. (2011). More than Measurement: The TAP system's lessons learn for designing better teacher evaluation systems. *National Institute for Excellence in Teaching* Retrieved on February 3, 2011, from http://www.tapsystem.org/publications/eval_lessons.pdf
- Johnson, S. M. (1989). Schoolwork and its reform. In J. Hannway & R. Crowson (Eds.), *The politics of reforming school administration: The 1988 yearbook of the politics of education administration* (pp. 95-112). Philadelphia: The Falmer Press.

- Lambert, L. (2003). Leadership redefined: An evocative context for teacher leadership. *School Leadership & Management*, 23(4), 421-430.
- Jordan, H. R., Mendro, R. and Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement: A preliminary report on research on teacher effectiveness*. National Evaluation Institute, Indianapolis, IN.
- Kane, T. J. and Cantrell, S. (2010). Learning about Teaching: Initial findings from the Measures of Effective Teaching Project. *Bill and Melinda Gates Foundation* Retrieved on December 21, 2010, from Learning about Teaching: Initial findings from the Measures of Effective Teaching Project
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011, June 1). Evaluating Teacher Effectiveness. . Retrieved May 26, 2014, from <http://educationnext.org/evaluating-teacher-effectiveness/>
- Klein, A. (2015, April 10). No Child Left Behind: An Overview. Retrieved April, 2017, from <http://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html>
- Lacireno-Paquet, N., Bocala, C., & Bailey, J. (2016). Relationship between school professional climate and teachers' satisfaction with the evaluation process.
- Lambert, L. (2003). Leadership capacity for lasting school improvement. Alexandria, VA: ASCD.
- Levenstein, R. (2016). 2014 My Voice, My School Survey Quality Profile.
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- Lustick, D. & Sykes, G. (2006). *National Board Certification as professional development: What are teachers learning?* Arlington: National Board for Professional Teaching Standards.
- Mathers, C., Oliva, M. and Laine, S. W. M. (2008). Improving Instruction Through Effective Teacher Evaluation: Options for states and districts. *National Comprehensive Center for Teacher Quality* Retrieved on November 11, 2010, from <http://www.tqsource.org/publications/February2008Brief.pdf>
- Marzano, R. J. (2007). [*The art and science of teaching: A comprehensive framework for effective instruction*](#). Alexandria, VA: ASCD.
- Marzano, R. J. (2009). Setting the record straight on "high yield" strategies. *Phi Delta Kappan*, 91(1), 30–37.
- Marzano, R. J. (2012). The Two Purposes of Teacher Evaluation. *Educational Leadership*, 70(3), 14-19. Retrieved from <http://www.ascd.org/publications/educational-leadership/nov12/vol70/num03/The-Two-Purposes-of-Teacher-Evaluation.aspx>

- Marzano, R. J., Frontier, T., & Livingston, D. (2014). *Effective supervision: Applying The Art and Science of Teaching*. Alexandria, VA: ASCD.
- Marzano Research Laboratory. (2010). *An observational protocol based on The Art and Science of Teaching*. Retrieved October 2010, from www.marzanoresearch.com/documents/MarzanosObservationalProtocol_Updated_Feb2010.pdf
- Margolis, J. (2008). What will keep today's teachers teaching? Looking for a hook as new career cycle emerges. *Teachers College Record*, 110(1), 160-194.
- Maryland's Race to the Top Application. (2014). Retrieved January, 2017, from <https://www2.ed.gov/programs/racetothetop/phase2-applications/maryland.pdf>
- Mead, S. (2012). Recent State Action on Teacher Effectiveness: What's in State Laws and Regulations?. *Bellwether Education Partners*.
- Merriam-Webster. (2004). *Merriam-Webster's collegiate dictionary*. Merriam-Webster.
- Millman, J., & Darling-Hammond, L. (1991). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Corwin Press.
- Minnici, A. The Mind Shift in Teacher Evaluation. *American Educator*, Spring 2014
- McCaffrey, D. F. and Lockwood, J. R. (2008). *Value-Added Models: Analytic Issues*. National Research Council and the National Academy of Education, Board on Testing and Accountability Workshop on Value-Added Modeling, Washington, DC. Retrieved on January 7, 2011 from http://www7.nationalacademies.org/bota/VAM_McCaffrey_Lockwood_Paper.pdf
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M. and Hamilton, L. (2003). Evaluating Value-Added Models for Teacher Accountability. *RAND Corporation* Retrieved on November 4, 2010, from http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf
- McCaffrey, D. F., Sass, T. R. and Lockwood, J. R. (2008). The Intertemoral Stability of Teacher Effect Estimates. Retrieved on November 2, 2010, from <http://mailer.fsu.edu/~tsass/Papers/Intertemporal%20Stability%20of%20Teacher%20Effects%2034.pdf>
- McGuinn, P. (2010). Creating Cover and Constructing Capacity: Assessing the Origins, Evolution, and Impact of Race to the Top. Education Stimulus Watch. Special Report 6. American Enterprise Institute for Public Policy Research.
- Meyers, T. (2012). An evaluation architect says teaching is hard, but assessing it shouldn't be. *Schoolbook*.

- Motel, L. (2010). Maryland's Race to the Top (RTTT). Retrieved January, 2017, from http://archives.marylandpublicschools.org/MSDE/programs/race_to_the_top/
- National Board for Professional Teaching Standards (2016). *Building to Accomplished Practice*. Retrieved February 2016, from <http://www.nbpts.org/atlas>.
- National Council on Teacher Quality. (2012). State of the states 2012: Teacher effectiveness policies. Washington, DC: Author. <http://eric.ed.gov/?id=ED536371>
- Park, S., Oliver, J. S., Johnson, T. S., Graham, P., & Oppong, N. K. (2007). Colleagues' roles in the professional development of teachers: Results from a research study of National Board certification. *Teaching and Teacher Education*, 23, 368-389.
- Peterson, P. L. and Comeaux, M. A. (March 20, 1990). Evaluating the Systems: Teachers' Perspectives on Teacher Evaluation. *Education Evaluation and Policy Analysis* 12(1).
- Pucella, T. J. (2011). The impact of National Board Certification on burnout levels in educators. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 84(2), 52-58.
- Race to the Top Resources | U.S. Department of Education. (2013, July 24). Retrieved April 19, 2017, from <http://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/index.html>
- Rogers, C. B. H., & Weems, D. M. (2010). Are US teachers making the grade?; A proposed framework for teacher evaluation and professional growth. *Management in Education*, 24(1), 19-24.
- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and teacher Education*, 23(2), 146-159.
- Ross, J. A., Hogaboam-Gray, A., & McDougall, D. (2002). Research on reform in mathematics education, 1993-2000. *Alberta Journal of Educational Research*, 48(2), 122.
- Roy, T., & Heflebower, T. (2012). Coaching classroom instruction. Solution Tree Press.
- Sass, T. R. (2008). The Stability of Value-Added Measures of Teacher Quality and Implications for Teacher Compensation Policy. *CALDER* Retrieved on November 2, 2010, from http://www.urban.org/uploadedpdf/1001266_stabilityofvalue.pdf
- Sawchuk, S. (2015) "New Studies Affirm Impact of Board-Certified Teachers." *Education Week*. Retrieved from: <http://www.edweek.org/ew/articles/2015/04/01/board-certified-teachers-more-effective-new-studies-affirm.html>

- Schochet, P. Z. and Hanley, C. S. (2010). Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains (NCEE 2010-4004). *National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education* Retrieved on November 3, 2010, from <http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf>
- Schon, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Silverman, S. (2017, March). WHAT WILL HAPPEN TO TEACHER EVALUATION Under the Every Student Succeeds Act? Retrieved April, 2017, from <https://thelinek12.com/what-will-happen-to-teacher-evaluation-under-the-every-student-succeeds-act/>
- Stigler, J. W. (2010). Needed: Fresh thinking on teacher accountability. *Education week*, 29(33), 30.
- Stiggins, R. (2014). *Defensible teacher evaluation: Student growth through classroom assessment*. Corwin Press.
- Stronge, J. H. (2006). Teacher evaluation and school improvement: Improving the educational landscape. In J. H. Stronge (Ed.) *Evaluating teaching: A guide to current thinking and best practice* (pp. 1–23, 2nd ed.). Thousand Oaks, CA: Corwin.
- Stronge, J. H. (2007). Qualities of effective teachers (2nd Edition). Chapter 4 *Planning and organizing for instruction*. Association for Supervision and Curriculum Development.
- Taylor, E. S., & Tyler, J. H. Can Teacher Evaluation Improve Teaching?. *Education Next*, 12. Retrieved May 26, 2014, from <http://educationnext.org/can-teacher-evaluation-improve-teaching>
- Taylor, E. S., & Tyler, J. H. (2011, March). The effect of evaluation on performance: evidence from longitudinal student achievement data of mid-career teachers. Retrieved May 26, 2014, from <http://www.nber.org/papers/w16877.pdf>
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *The American Economic Review*, 3628-3651.
- Teacher Evaluation. (2016). Retrieved March, 2017, from <https://www.nea.org/assets/docs/Backrounder-Teacher%20Evaluation.pdf>
- Teacher Evaluation and Support Systems: A Roadmap for Improvement [Scholarly project]. (2016, March). Retrieved April, 2016, from <https://www.aspeninstitute.org/publications/teacher-evaluation-and-support-systems-a-roadmap-for-improvement/>
- Teacher Performance Assessment System (TPAS), St. Mary's County Public Schools. (2010, August). Retrieved March, 2015, from http://www.smcps.org/files/TLPD/TL/TPAS_2010.pdf

- The ABC's of ESEA and No Child Left Behind. (n.d.). Retrieved April, 2017, from <http://educationpost.org/issues/taking-responsibility/esea-reauthorization/abcs-esea-child-left-behind/>
- Thessin, R. A., & Starr, J. P. (2011). Supporting the growth of effective professional learning communities' districtwide: teachers do not magically know how to work with colleagues; districts must support and lead that work if PLCs are to live up to their potential. *Phi Delta Kappan*, 92(6), 48.
- Thorpe, R. (2014). Sustaining the teaching profession. *New England Journal of Public Policy*, 26(1), 5.
- Toch, T. and Rothman, R. (2008). Rush to Judgment: Teacher Evaluation in Public Education. *Education Sector* Retrieved on January 11, 2011, from http://www.educationsector.org/usr_doc/RushToJudgment_ES_Jan08.pdf
- Tucker, P. D., Stronge, J. H. and Gareis, C. R. (2002). *Handbook on Teacher Portfolios for Evaluation and Professional Development* Larchmont, NY: Eye on Education.
- Tucker, P. D., & Stronge, J. H. (2005). Linking teacher evaluation and student learning. ASCD.
- Tyler, John (2011). "Designing High Quality Evaluation Systems for High School Teachers" (Washington: Center for American Progress, 2011), available at http://www.americanprogress.org/wp-content/uploads/issues/2011/11/pdf/high_school_evaluation.pdf
- Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation. Research Report*. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.
- Sato, M., Chung, R. R., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45(3), 669-700. doi: 10.3102/0002831208316955.
- Shulman, L. S., & Shulman, J. H. (2004). How and what teachers learn: A shifting perspective. *Journal of Curriculum Studies*, 36(2), 257-271.
- Uhlenbeck, A. M., Verloop, N. and Beijaard, D. (2002). Requirements for an Assessment Procedure for Beginning Teachers: Implications from Recent Theories on Teaching and Assessment. *Teachers College Record* 104(2), 242-272 Retrieved on March 23, 2011, from <http://www.tcrecord.org/Content.asp?ContentID=10828>
- U.S. Department of Education. (2010). Race to the Top application for initial funding. Washington, DC: Office of Elementary and Secondary Education. Retrieved from www2.ed.gov/programs/racetothetop/application.doc

- Valli, L., van zee, E. H., Rennert-Ariev, P., Mikeska, J., Catlett-Muhammad, S., & Roy, P. (2006). Initiating and sustaining a culture of inquiry in a teacher leadership program. *Teacher Education Quarterly*, 33(3), 97-114. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=21639616&site=ehost-live>
- Weingarten, R. (2011, February 24). Toward a True Development and Evaluation System. Retrieved March, 2011, from http://www.aft.org/sites/default/files/wysiwyg/sp_weingarten022411.pdf
- Watson, R., & Manning, A. (2008). Factors influencing the transformation of new teaching approaches from a programme of professional development to the classroom. *International Journal of Science Education*, 30(5), 689-709. doi: 10.1080/09500690701854881.
- Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the United States and abroad. Dallas, TX: National Staff Development Council.
- Whitehurst, G., Chingos, M. M., & Lindquist, K. (2015). Getting Classroom Observations Right. *Education Next*, 15(1), 62-68.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project. Retrieved June 2014, from <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>
- Wenner, J. A., & Campbell, T. (2016). The theoretical and empirical basis of teacher leadership: A review of the literature. *Review of Educational Research*. Doi: 10.3102/0034654316653478.
- Wise, A. E. (1984). *Case Studies for Teacher Evaluation: A Study of Effective Practices*.

