2018

Teachers' Perceptions of Value-added Models for Use in Annual Performance Appraisals: a Case Study

Theresa M. Pascual
Bethel University

Teachers' Perceptions of Value-added Models for use in Annual Performance Appraisals: A Case Study

by
Theresa M Pascual

A dissertation submitted to the faculty of Bethel University
in partial fulfillment of the requirements for the degree of
Doctor of Education

St. Paul, MN
2018

Approved by:

Advisor: Michael Lindstrom, Ed.D.

Reader: Adam Wyse, Ph.D.

Reader: Barry Sullivan, Ph.D.

Abstract

Motivating and retaining teachers is a critical element of the educational process. In 2011, Florida law required school administrators to implement a Valued-Added Model (VAM) for teacher evaluation, compensation, and retention decisions. However, the Florida VAM includes complex calculations to measure student learning progress. This qualitative case study was to explore teachers' professional and personal perceptions of the VAM to determine teachers' performance, compensation, and personnel consequences. Data collection involved semi-structured interviews with 12 teachers employed at a single Florida school district. Participants held a negative view on the use of VAM for teacher evaluations and merit pay determination. A majority of teachers had a limited understanding of inputs and formulas used in the VAM calculations and few understood how VAM scores related to teaching practices. None of the teachers could accurately describe how the scores were used to calculate the VAM score. Teachers resented that factors beyond their control affected VAM scores, such as the proportion of students with learning disabilities in their classroom, and students' home life problems based on socioeconomic variables. School administrators need to invest the time and resources necessary to educate teachers on the VAM system and reflect their input in the calculation.

Acknowledgements

My doctoral journey would not have been possible had it not been for the love and support of my husband, Raul. My dear husband spent many evenings, weekends, and summers caring for our children so that I could have the time to complete this rewarding venture. I know I gave up many trips to theme parks and beaches in order to read books and sit in front of a computer to work on this dissertation. It was not easy to complete this dissertation and my principal internship while working full-time, but my husband's support throughout this whole process is what helped me to press on in nurturing both my mind and our children to the best of my ability. Without a doubt, having my supportive husband has been integral to my success.

I also would like to dedicate this dissertation to my father who went to heaven before I could finish. During my time of sorrow, my father's spirit gave me the strength to keep going on this journey and never give up. I know he would be very proud of my accomplishment.

I am thankful for my cohort who provided constant motivation to help me realize, I too could do this, and complete my degree.

I am thankful to Dr. Michael Lindstrom for his commitment to me as advisor. His inspiring words and emails, as well as the informative documents he would send me, helped inspire me to complete my dissertation.

Finally, I am thankful to God for helping me find Bethel University as it has been a tremendous blessing in my life. I am so grateful for the biblical worldview education I received and that I was able to be educated from the comfort of my home in Florida with minimal time away from my family in Minnesota while on campus. The flexibility of being able to complete courses online was one of the greatest benefits to me, and my family; without that option, I am not confident that this doctoral degree would have been attainable.

Table of Contents

List of Tables

Chapter I: Introduction

Introduction

Are teachers being "Vamboozled"? Vamboozled is a term used to characterize teachers' experience in America's public schools since the implementation of the value-added model (VAM) for teacher evaluations (Amrein-Beardsley, 2014). Educators nationwide are reassessing the value and fairness of the VAM. It has become abundantly clear that maximizing teacher effectiveness has become an increasingly important nationwide education priority, fueled not just by growing awareness of compelling research, but also by a new federal emphasis on the teacher performance due to Race to the Top Funding (R2T; LaVenia, Cohen-Vogel, & Lang, 2014).

In an effort to qualify for R2T, the State of Florida's legislature passed Senate Bill 736, the Student Success Act (SSA), which revised the procedures for teachers' annual evaluations (State Impact, 2014). The SSA involved a rigorous annual evaluation that emphasized students' performance on standardized exams to measure teacher performance and recommend merit pay adjustments, referred to as a Value-Added Model (VAM). The model incorporates student standardized test scores into the equation to determine teacher performance excluding other relevant factors, such as student hunger, sickness, stress, or major life events. The output from the VAM equals 50% of a teacher's evaluation while the other 50% is classroom observations. VAM plus classroom observation holds teachers accountable for student learning outcomes, and identifies effective and ineffective teachers if necessary.

Before the SSA, Florida had the "Florida A+ Plan" which would grade the schools based on the student learning gains (LaVenia et al., 2014). In 1999, Jeb Bush, who was governor of the State of Florida, implemented reforms that emphasized school choice, annual tests, grading of schools and districts on an A- through F-based grading system, required illiterate children to

repeat third grade, and gave performance bonuses to teachers. This reform also stated that any student in a public school, rated F twice in four years, could get a voucher to move to a different public or private school. This A+ plan assigned accountability of student learning growth to an entire school and school district.

In the new Florida SSA evaluation system, all variance in student performance on standardized tests is attributed to teacher performance (State Impact, 2014). Florida created a VAM that uses the current year and two previous years of student test scores to determine if a teacher is effective or ineffective. The VAM is a predictive model that uses students' prior performance on standardized tests to estimate future performance. Predictor variables in the VAM formula utilized to measure teacher effectiveness include:

1. Number of subject-relevant courses in which the student is enrolled.
2. Two prior years of achievement scores
3. Disabilities (SWD) status
4. English language learner (ELL) status
5. Gifted status
6. Attendance
7. Mobility (number of transitions).
8. Difference from modal age in grade as an indicator of retention.
9. Class size: A continuous measure counting the number of students linked to teacher.
10. Homogeneity of entering test scores in the class (Sass, Semykina, & Harris, 2014).

The VAM does not predict a student's unique, individual experience, which may or may not affect the validity of the test and does not include factors such as poverty (LaVenia et al., 2014). The Florida Department of Education (FLDOE) excluded socioeconomic factors. The

rationale behind this is regardless of whether or not a student is homeless or poor, with a good teacher, their scores will improve. A well-founded evaluation requires that one must be able to interpret the results as an accurate reflection of the intent of the assessment. Validity is a characteristic of inferences drawn from the results of teacher assessments (LaVenia et al., 2014). A current controversy among educators in the state of Florida is whether the Florida State Assessment (FSA) test is valid. In a statement given on their website the FLDOE affirmed that the FSA was indeed valid. The new FSA tests were used statewide during the spring of 2015. Schools encountered problems with the administration, the training, and the delivery of the tests. Since teachers are being held accountable for students' learning growth with the VAM, they need to have a basic understanding of it, and how their students are being measured on the new FSA.

On July 1, 2017, House Bill 7069, a 274-page $419 million dollar measure came into effect to reform Florida's K-12 schools with dozens of changes. In the bill, the Florida VAM may be used for teacher evaluations, but is no longer required by state law. Districts are allowed to determine how learner growth is measured, but teacher evaluations must factor in student performance in some way (Florida Senate, 2018). The bill also mandates that effective teachers can receive up to an $800 bonus and $1200 for highly effective teachers. These bonuses will have no permanent effect on salary or will count towards retirement. If more teachers qualify than expected, then bonuses will be pro-rated. In addition, if a teacher wants to sue over a VAM score, they will have to sue their district, and not the state of Florida (Florida Senate, 2018). The Florida school district utilized for this study is utilizing VAM as 35% of a teacher's evaluation for teachers of elementary, English Language Arts, and those that use the Algebra 1 end of course exam. For classroom teachers of courses for which there are no statewide assessments, districts may use measureable learning targets approved by the principal.

Other states' experiences. Texas, California, and Tennessee, among others, adopted teacher evaluation schemes using VAMs (Sass et al., 2014). Each is similar in that they are purely statistical measures based solely on standardized student test scores. Classroom observations, other measures of student learning, and interviews with students, teachers or administrators are gone. Statistically speaking, all of these models have a fundamental research design flaw in that there are no random pairings among students and teachers to infer causal attributions.

The result of applying purely statistical models resulted in obvious conflicts between the "calculated" performance and observations (Sass et al., 2014). In fact, applying different algorithms created different results for the same teacher. The Educational Value-Added Assessment System (EVAAS), developed by William Sanders and his associates for use in Tennessee, is the most widely used VAM. EVAAS, implemented in 1993, was adopted by districts in seven other states (Sanders & Horn, 1994). The Dallas Value-Added Accountability System (DVAAS) is a widely cited alternative to the EVAAS and has been employed by the Dallas school system for a number of years. DVAAS uses a student improvement on standardized test improvement as the value-added criterion to identify effective teachers, and those in need of support. DVAAS differs from EVAAS in four critical ways. First, DVAAS use student-level characteristics, such as socio-economic status, to adjust student test scores prior to analysis. Second, the model includes test scores in adjacent grades (as opposed to combining several grades). Third, it adjusts test scores to reflect a more general structural connection. Finally, the model includes not only student achievement but also other factors that reflect student achievement. Another VAM alternative, the rate of expected academic change (REACH), created by Doran and Izumi (2004) was in use in California. REACH test-based criterion

measures student progress toward a proficiency standard. Thus, each student's growth is measured against a goal rather than against the growth of other students. Such a VAM could also be used to evaluate teacher effectiveness, but do not suggest a particular model for evaluating the teachers' contribution to academic achievement.

Statement of the Problem

Beginning with the 2014-2015 school year, a school principal's judgment and the VAM score each contribute 50% to teacher evaluation and are used for compensation and retention decisions (State Impact, 2014). VAM quantifies teachers' performances based on expected, versus actual, performance of students' standardized tests. VAM involves a complex statistical formula to predict students' future performance based on their past scores, disabilities, gifted status, and movement from school to school, as well as other factors.

Since teaching quality plays such an important role in student learning and academic progress, identifying effective and ineffective teachers is of critical importance. A successful teacher evaluation system is key to advancing student learning. Well-designed and implemented teacher evaluations should identify and evaluate the instructional strategies, professional behaviors, and delivery of content knowledge that affect student learning (State Impact, 2014). The use of the VAM linked to teacher effectiveness is a cause for concern as Braun (2005) noted, "causal attributions cannot be confidently made about the quality of teaching due to the lack of randomization – no matter how complex the statistical model is and how sophisticated the method of analysis is" (p. 10). There could be many other unmeasured attributes associated with the results from VAM models, which, when used in high-stakes situations, can bring unintended negative consequences. Therefore, the results from student learning outcomes should be

properly used to improve teaching and learning and inform decision making, rather than only with high-stakes accountability purposes (Braun, 2005).

SB 736 legislation requires that teachers in Florida be evaluated by their students' scores. The overall problem is that the VAM is a complex statistical formula that teachers and administrators struggle to understand. When looking at the VAM model, many teachers admit to having no idea how it relates to what they or their students are doing in the classroom. Teachers need to understand how they will be evaluated and for what factors affecting student-learning growth they will be held accountable.

Purpose

The purpose of this qualitative case study was to explore teachers' professional and personal perceptions of the VAM to determine teachers' performance, compensation, and personnel consequences. While teacher evaluations were originally designed to serve summative (accountability) and formative (improvement) functions, the shift toward accountability may hamper teachers' professional development (Donaldson & Johnson, 2010). Teacher evaluations are frequently the only annual opportunities for principals to observe, deliver constructive feedback, and plan for professional development opportunities. When teacher evaluations are exclusively used as a means of delivering consequences, the formative function is lost. Absent formative feedback, there is no mechanism to promote teacher quality improvement (Sass et al., 2014).

The study is relevant in light of the case, *Houston Federation of Teachers et al. v. Houston ISD* filed in May 2014, brought by seven Houston teachers along with the Houston Federation of Teachers to end the EVASS (Houston Federation of Teachers, 2017). Plaintiffs claimed that VAM reduced education to a test score did not improve teaching or learning, and

ruined teachers' careers when they were incorrectly terminated. The value-added methodology was used to make decisions about teacher evaluation, bonuses, and termination.

Because of the lawsuit, the EVAAS was eliminated in May 2017. The judge concluded that HISD teachers "have no meaningful way to ensure correct calculation of their EVAAS scores, and a result are unfairly subject to mistaken deprivation of constitutionally protected property interests in their jobs" (p. 18). HISD agreed never to use the value-added measures, including EVAAS scores and agreed to create an instructional consultation panel with representatives from the district and the Houston Federation of Teachers to discuss and make recommendations on the district's teacher appraisal process.

Definition of Terms

Student Success Act. The Student Success Act (SSA), or Senate Bill 736, passed the Florida State Legislature in 2011, revised the procedures for teachers' annual evaluations to include the VAM (State Impact, 2014).

Value-added Model. The value-added model (VAM) refers to an algebraic formula used to predict students' future performance on standardized based primarily on a student's most recent two-year performance on standardized test. The FLDOE (2017) requires teachers' annual performance appraisal include VAM results, weighted at 35%, and classroom observations, weighted at 35%.

Research Questions

The following research questions helped to guide the study:

RQ1. Do teachers believe the use of student standardized test performance should be used in the VAM to determine 35% of their performance appraisal?

RQ2. To what degree do teachers perceive that data generated from the VAM and classroom observations improve teaching practices?

RQ3. To what degree do teachers perceive that performance appraisal using the VAM, weighted at 35%, should inform compensation and personnel decisions?

Significance of the Study

In the era of increased accountability, teachers are held responsible for unsatisfactory educational outcomes, including poor student performance and high dropout rates (State Impact, 2014). Evaluation is becoming more and more important as a means of determining a teacher's effectiveness, retention, and pay for performance. Evidence is needed regarding the validity of using the VAM to hold teachers responsible for both the problem and the solution to unsatisfactory outcomes, since it determines a teacher's future. It is also important to understand a teacher's knowledge of the VAM and if the VAM reflects classroom performance accurately. This study looks into the lived experiences of teachers and how they understand VAM scores, to determine if errors exist in their actual performance appraisals or in their understanding of the VAM. The findings of this study could contribute to the ever-growing research and the reliability and validity of VAM being used in teacher evaluation systems.

Chapter II: Review of Literature

Using the value-added model (VAM) score in a teacher's evaluation affects teachers in various ways. Research for this study showed what teachers knew and how they felt about the VAM being used for their evaluations, and ascertained if they were aware of what their evaluations meant, how they were calculated, and how they perceived the effects on them personally, as well as on their careers. The thoughts and attitudes of teachers towards their evaluations were explored. Various studies and research from educational experts on understanding the VAM and how it affects teachers was explored and included in the literature review.

Purpose and History of Teacher Evaluation Systems

Teacher evaluation systems are designed to promote teaching practice improvement and support continuous professional development (Donaldson & Johnson, 2010). Teacher evaluations play a critical role in nurturing teacher instructional ability and are often the only annual coaching opportunity. Teacher evaluation systems "provide a forum, structure, and plan for teachers and evaluators on which to reflect, change, and assess professional practice" and provide feedback (Feeney, 2007, p. 191). Feedback refers to communication between an evaluator and teacher as a technique to improve performance, make modifications, correct errors, and create a plan for continuing professional development. A high quality, actionable evaluation system for teacher improvement (Danielson, 2011):

1. Is based on recent objective observations,
2. Is provided to promote specific improvements in quality,
3. Includes timely suggestions for improvement,
4. Promotes reflective inquiry toward professional development, and

5. Reflects teachers' current developmental tasks.

Teacher evaluation feedback is most effective promptly after observation in a face-to-face session. Immediate feedback sessions are more valuable for behavior modification than a five-page memo a month later.

Teacher evaluations serve two simultaneous functions: summative (accountability) and formative (improvement) (Donaldson & Johnson, 2010). The public and policymakers generally perceive that teacher evaluations are exclusively a tool for ensuring quality instruction, which partially explains the evolution toward the use of student achievement to evaluate teachers (Danielson, 2011). However, since teacher evaluations are frequently the only opportunity for principals to provide constructive criticism and to coach teachers, a balance is needed between summative and formative feedback. When teacher evaluations are used as a means of delivering consequences, the formative function is lost. Absent formative feedback, there is no mechanism to promote teacher quality improvement (Sass et al., 2014).

Healthy feedback and fair teacher evaluations are the core of professional development and are critical for teachers' professional development (Sass et al., 2014). Principals historically addressed both the formative and summative needs using annual observations and performance reviews (Danielson, 2011). Teacher evaluation systems, such as VAMs, used exclusively as a job performance lose the capability to improve teacher performance. Standards-based evaluation systems that rely on administrators acting as judges are sustainable models for teacher professional development (Danielson, 2011).

Teachers engaged in a program of continuous professional development form the backbone of the educational process. "Teaching is the essence of education, and there is almost universal agreement among researchers that teachers have an outsized impact on student

performance" (Weisberg et al., 2009, p. 9). Development of appropriate evaluation systems to improve American education is the responsibility of school administrators. The goal of the teacher evaluation process is to ensure high quality education inside the classrooms to drive student achievement increases (Weems & Rogers, 2010).

Three significant educational reforms occurred in the U.S. since the publication in 1983 of A *Nation at Risk* (Weems & Rogers, 2010). The first educational reform involved increasing students' volume of academic work required for each class. The second phase involved the use of standardized tests to measure students' academic progress. The third educational reform was the formalized connection between students' academic achievement and teacher evaluations (Sass et al., 2014). While the processes for determining teacher performance through observation remain unchanged over the past half-century, the purpose and philosophy of teacher evaluation evolved (Weems & Rogers, 2010). Teacher evaluations during the 1970s involved observation of best practices teaching behaviors, and direct observation was the method for data collection. Beginning in the 1980s, education reform adopted teacher evaluations as a tool to improve student achievement (Weems & Rogers, 2010). Teacher evaluation expanded to include standardized test performance.

No Child Left Behind (NCLB) was implemented in 2001 where legislation placed an emphasis on student achievement and teacher quality. The quality of school staff along with student achievement was a key provision of NCLB. States were also required to be more involved with the teacher evaluation process (Barton, 2010). NCLB mandated that each school meet adequate yearly progress (AYP) targets based on standardized test scores and certain subcategories for at-risk children, such as low socio-economic status (SES) students. To meet

the higher proficiency goals, NCLB expanded oversight and mandated that only highly qualified teachers, defined by one's academic attainment, teach in classrooms (Weems & Rogers, 2010).

The U.S. Department of Education created the *Race to the Top* (R2T) grant program in 2009 to promote teacher quality improvement (U.S. Department of Education, 2009). States competed for R2T funds to improve education quality based on standardized test scores. R2T required new teacher evaluation systems that emphasized student achievement data as an input (FLDOE, 2017). R2T led many states to adopt VAM-related evaluation systems and imposed higher educational and professional development standards in order to satisfy the requirements to receive the R2T funding. Since R2T began, the number of states with annual evaluation of teachers based on student achievement scores increased from 15 to 23 (FLDOE, 2017). While not all states pursued R2T funding, those that did had to make statutory changes to meet the application criteria for R2T. NCLB, coupled with R2T, have permanently made student achievement a criterion for teacher evaluations in many states.

Teacher Evaluations in Florida

The Florida Statute (Florida Statute, 2011) indicates that teacher evaluation is:

> For the purpose of increasing student academic performance by improving the quality of instructional, administrative, and supervisory services in the public schools of the state, the district school superintendent shall establish procedures for evaluating the performance of duties and responsibilities of all instructional, administrative, and supervisory personnel employed by the school district (p. 1102).

While the state statutes indicate that evaluations are for the purpose of improving instruction, the current evaluation policy in Florida is not only aimed towards measuring the effectiveness of each teacher, but also categorizing and ranking teachers, rewarding those at the

top, and firing those at the bottom. The core of education is teaching and learning, and the teaching-learning connection works best when effective teachers work with students every day. Educational reform cannot succeed without capable, high quality, competent teachers in the classrooms (Stronge & Tucker, 2003).

The two most frequently cited purposes of personnel evaluation are accountability and professional growth (McGaghie, 1991). Accountability reflects the need for determining teachers' competence in order to assure that services delivered are safe and effective (McGaghie, 1991). Performance improvement reflects the need for professional growth and development of the individual teacher and is considered to be formative in nature.

Stronge and Tucker (2003) stated that comprehensive teacher evaluation systems should serve two purposes. First, it should be accountability-oriented, contributing to the personal goals of the teacher and to the mission of the program, the school, and the total educational organization, and should provide a fair measure of accountability of performance. Second, it should be improvement-oriented, contributing to the personal and professional development needs of the individual teacher as well as improvement within the school.

Political Issues that Affect Teacher Evaluations

A major obstacle to effective teacher evaluation systems can be the influence of politics. The process is described as both emotionally laden and politically challenging (Stronge & Tucker, 1999). The stakeholders involved in the development of the new evaluation system must buy into the new system. These stakeholders' expectations often conflict when deciding what is good practice and effective reform. However, the input and support of these groups is an important aspect to gaining political support for new evaluation systems (Stronge & Tucker, 1999). They have differing views on issues related to both improvement and accountability.

A controversial issue relative to teacher evaluation is merit pay. Rewarding teachers based on test scores and using test scores for evaluation have become important issues (Millman, 1997; Stronge & Tucker, 2003). Ramirez (2001) said that teacher compensation should be based on experience and degree of advancement. Teachers are continually gaining valuable on-the-job-experience and the skills to improve their pedagogy (Ramirez, 2001). While the experience and degree of attainment approach was widely used for many years, the chronic achievement gap between the U.S. and other developed nations and the adoption of NCLB resulted in a business model approach where compensation is based on quality of work (Hechinger Report, 2011).

From 2007 to 2010, the New York City Department of Education and the United Federation of Teachers offered bonuses to a random sample of the city's high-needs public schools. The RAND Corporation was hired to study the program's results. According to RAND, extra pay did not necessarily improve student achievement because conditions needed to motivate the teaching staff were not achieved e.g., understanding, buy-in for the bonus criteria) and because of pressure to gain a high level of accountability (Hechinger Report, 2011). While one might object that the standardized New York State exams used to evaluate the bonuses were a poor and unreliable measure of student achievement, the report's other findings were that the bonuses were seen as weak motivation that did not change educator behavior or practice. A teacher's primary motivation in entering the field of education is not economic gain, but to make a difference in the lives of the young people (Hechinger Report, 2011).

Maslow's (1943) study of human motivation also informs policymakers with regard to motivation of educators. Maslow based his theory on a hierarchy of needs, to which all humans respond. This hierarchy ascends from basic, to complex needs like money, benefits, and job security, which appear at the lower end of the hierarchy (Maslow, 1943).

Another less publicized political issue that relates to teacher evaluation is the issue of how to implement a new teacher evaluation system without creating a political problem in the district or the school. Stronge and Tucker (1999) conducted case studies that yielded recommendations on how to effectively implement standards-based teacher evaluation systems. Regardless of how well a program is designed, it is only as effective as the people who implement it (Stronge & Tucker, 1999). According to Stronge and Tucker (2003), teachers are the most important factor in schools. In order to have dramatic improvements in all students' preparation for college and careers, states will need to implement well thought human capital strategies that put the right teachers in the right schools teaching the right subject matter. Stronge (2007) highlighted the following commonalities of effective teachers: strong classroom management, good delivery of instruction, and consistent monitoring of student progress. A teacher's verbal ability, educational coursework, teacher certification, content knowledge, and teaching experience have an impact on teacher effectiveness (Stronge, 2007). Teachers must be caring, fair, respectful, promote enthusiasm, and motivate learning (Stronge, 2007).

Perceptions of Feedback from Teacher Evaluations

Teacher evaluation systems involve groups of people and human behavior. Some resources examine the psychology of teacher evaluation and the perceptions of the feedback from these evaluations. Research from Conley, Muncey, and You (2005) revealed mixed opinions and levels of satisfaction when standards-based teacher evaluation was implemented due to role ambiguity and work criteria autonomy.

Milanowski (2005) examined the problem of the principal's split role of evaluator and mentor and concluded there is little impact with one supervisor filling both roles. The program design described in the study provided new teachers with a single mentor or an administrator and

a mentor to assess teachers' progress. Mentors evaluated and provided feedback for teacher interns in the first year of induction along with administrators. Results of the study did not show significant differences in teachers' evaluation based on roles. However, the study suggests the quality and consistency of the assistance teachers receive is important. Bouchamma (2005) surveyed over 300 teachers in Canada regarding who should provide their supervision. They found teachers preferred supervision by the school principal to self-evaluation, peer evaluation, and student evaluation, with the least preferred being no evaluation.

One effect of state-mandated assessments is teachers' perceived feelings of test-related pressure to improve student achievement (Papay & Johnson, 2012). Research indicates that teachers seem to share similar perceptions of test-related pressure. Several studies have established that teachers throughout the United States are not opposed to accountability (McGaghie, 1991; Stronge & Tucker, 2003). E-mail data from teachers in Texas showed they were not against accountability, yet contended that classroom assessment served as a more valuable tool in informing their instruction rather than high-stakes tests (Booher-Jennings, 2005). In another study, educators surveyed in Texas indicated that teachers should be held accountable for their teaching, but did not believe their state's high-stakes achievement test was an accurate measure of students' learning (Reese, Gordon, & Price, 2004). Similarly, interview data from teachers in Illinois indicated that although most agreed with being held accountable for their students' knowledge of state standards, they disagreed with the amount of emphasis placed on high-stakes testing (Stitzlein, Feinberg, Greene, & Miron, 2007).

A national study of elementary through high school educators showed that teachers working in schools with higher levels of poverty felt more pressure to raise students' test scores (Moon, Brighton, Jarvis, & Hall, 2007). Interview data from teachers working in an

impoverished urban Texas neighborhood conveyed their perceptions of test-related pressure to increase students' scores as one of the most significant effects of testing, whereby low-performing students were perceived as liabilities (Booher-Jennings, 2005). Hoffman, Assaf, and Paris (2001) documented teachers' frustrations over being compared with educators of middle-class students when held accountable for their economically disadvantaged students' test scores. Teachers agree that test-related pressures have led good educators to flee the teaching profession altogether.

Teachers Helping Teachers as Peer Assessors Improve Evaluations

Historically, teacher evaluations by principals showed significant and persistent bias in several ways (Aldeman & Chuong, 2014). First, teachers of gifted, or high achieving classes received higher performance appraisals than teachers in low achieving classes. Second, teacher performance appraisals were higher for classes with higher initial standardized test scores than those with lower initial scores. Third, teacher performance appraisals are skewed toward excellent performance, rather than a broad spectrum of quality as is typical in work environments. Therefore, the principal observation appraisal approach results in generally inflated scores. A variety of approaches is employed to address inflated appraisals. In Georgia, for example, rather than having 90% of teachers receiving the highest possible rating, under the new system only 20% of all teachers in any school receive the highest rating of "exemplary" (Aldeman & Chuong, 2014). In Louisiana, instead of evaluating 99% of teachers as "satisfactory," one-third of all teachers are designated as top performers, or "highly effective," under a new evaluation system (Aldeman & Chuong, 2014).

There is promising evidence that teachers serving as assessors in teacher evaluations are positive in improving practice and addressing rater bias. Peer assistance programs offer helpful

support to new and veteran teachers in need of improving their skills or knowledge. Most peer review programs have some form of peer assistance in place, thus connecting formative and summative aspects of teacher evaluation (Papay & Johnson, 2012). In school districts such as Toledo, Ohio, and Rochester, New York, where peer review programs have been implemented, the percentage of teachers who have received less-than-satisfactory evaluations and, thus, additional assistance and training, has increased dramatically over traditional administrator-only evaluations (from 0.1% to 8%).

Significantly higher percentages of first year teachers have been identified as needing assistance or as not satisfactory through peer-review. Interestingly, anecdotal accounts suggest that new teachers need and welcome assistance from more experienced colleagues, even when those colleagues render a negative evaluation. New teachers in Columbus, Ohio remain in their jobs longer than in typical urban districts lacking these programs with 80% remaining on the job after five years later (Papay & Johnson, 2012). There is evidence that peer review programs even help satisfactory teachers become better.

Research by Johnson, Papay, Fiarman, Munger, and Qazilbash (2010) funded by the Bill and Melinda Gates Foundation, examined seven districts utilizing a Peer Assistance and Review program (PAR) to see if peer review can improve teacher evaluations. A peer reviewer provides a teacher with subject-matter expertise that a principal may lack. However, research also suggests that without support from the administration, peer reviewers' advice and judgement lacks credibility.

Districts with fully implemented programs retained more novice teachers and dismissed more underperforming teachers, both tenured and non-tenured, than did comparable districts (Papay & Johnson, 2012). Research in the seven districts involved suggested peer review can

work well if the key components of (a) an open and rigorous selection, (b) clear performance guidelines, (c) explicit instructional standards, (d) ongoing training, and (e) effective supervision, are in place (Johnson et al., 2010). The researchers also note that the greatest benefit of consulting teachers is that they not only provide evaluation, but support as well.

Costs of VAM and Teacher Evaluations

Hillsborough County Schools in Florida were awarded up to $100 million in grant money from the Bill and Melinda Gates Foundation in November 2009 for utilizing the Empowering Effective Teachers (EET) evaluation system that placed peer reviews as 30% of the overall score. The peer review was one part of the evaluation system. A principal review comprised another 30%, while student learning gains from the VAM accounted for the final 40%. The goal of the EET was to make school better for children, particularly poor and minority children. The 2010 archives of Hillsborough schools "Empowering Teachers" webpage include enthusiastic responses regarding the newly-acquired $100 million dollar Gates grant. The premise behind the Gates name was to help Hillsborough County Schools be a model for the nation. The Gates-funded program went beyond the district's ability financially, creating a new bureaucracy of mentors and peer evaluators that did not work with students. Six years later the district's surplus significantly dwindled, risking its ability to borrow money (Herlihy et al., 2014). After investing in an elaborate system to improve evaluations, the new superintendent retreated from the EET model. According to various local news reports, The Gates Foundation also refused to pay the remaining $20 million grant money due to Hillsborough County Schools not following through with their promise to fire the 700 lowest ranking teachers. FLDOE data indicates that Hillsborough's graduation rate still lags behind other large school districts in the state. In

addition, racial and economic achievement gaps remained pronounced, especially in middle school.

A 2013 report by Rand Education and American Institutes for Research evaluated how much money school districts needed to implement the teacher evaluation system, which included Hillsborough County Schools. The report indicated that from 2009 to 2012 Hillsborough County Schools' spent $128.00 per pupil, or 24.8 million on VAM (American Institutes for Research, 2013). Rand estimated that 34.5% of total expenditures were made on the entire effective teaching initiative between 2009 and 2012. Rand also reported that in the 2010-2011 academic school year, less than 1% of the total district operating expenditures was used for teacher compensation expenditures.

Research on VAM Methods to Evaluate Teachers

Haertel (2013) published a detailed report on the lack of reliability using student test scores to evaluate teachers. Haertel (2013) concluded that VAM scores should not be an included substantial factor in teacher personnel decisions. The information provided was simply not good enough to use. Much more serious, the scores may be systematically biased for some teachers, and against others. Teacher VAM scores could easily have additional negative consequences for children's education. These consequences include (a) increased pressure to teach to the test, (b) more competition and less cooperation among the teachers within a school, and (c) resentment or avoidance of students who do not score well. In the most successful schools, teachers work together effectively. Placing teachers in competition with one another for bonuses or future employment is not beneficial to the students or the peer and mentoring relationships that are supposed to support new teachers (Haertel, 2013).

VAMs are complex statistical models requiring high-level expertise and awareness of their assumptions and limitations, especially when used for high-stakes purposes (Chetty, Friedman, & Rockoff, 2014). Recent studies indicate that few, if any, state education departments have the requisite statistical and technical expertise to use VAM models appropriately (Herlihy et al., 2014; McGuinn, 2012). In Florida, for example, teachers are partially rated based on school-wide achievement improvement, thereby evaluating teachers based on students they never taught. A group of teachers and their unions in Northern Florida filed a lawsuit in April 2013 challenging the constitutionality of Florida's teacher evaluation system using school-wide standardized test scores (Herlihy et al., 2014). The teachers and unions claimed that using test scores of students they do not teach or from subjects they do not teach is unfair and violates the Equal Protection and Due Process Clause of the Constitution.

The court held the plaintiffs' Due Process and Equal Protection claims fail because the evaluation policies pass a rational basis review. The lawsuit states, "While the Florida Comprehension Assessment Test (FCAT) VAM may not be the best method for achieving this goal, it is still rational to think that the challenged evaluation procedures would advance the government's stated purpose" (U.S. Court of Appeals, 2015, pp. 12-13). The court reasoned that Florida officials could reasonably believe that "a teacher can improve student performance through his or her presence in a school and the FCAT VAM can measure those school-wide performance improvements, even if the model was not designed to do so" (U.S. Court of Appeals, 2015, p. 14). Furthermore, the court believes that a teacher can improve student performance across subjects and that the FCAT VAM can measure school-wide performance improvements. Thus, the evaluation system measuring teacher performance by looking at other subjects, or school wide performance, could rationally lead to the improvement of students' test

scores, which was the stated purpose of the Student Success Act. Additionally, the current evaluation system gives teachers incentive to pursue school improvements, which can improve student performance (U.S. Court of Appeals, 2015).

VAMs are based on standardized tests and do not directly measure teacher contributions toward student outcomes and measure correlation, not causation (ASA, 2014). That means that the rise or fall of student test scores attributed to the teacher may be attributable to unmeasured factors not under the teacher's control. The VAM rating of teachers is so unstable that it may change if the same students complete a different test. Most VAM studies found that teachers account for roughly 1-14% of the variability in test scores; the majority of opportunities for quality improvement exist at the system-level conditions (American Statistical Association [ASA], 2014). Ranking teachers by their VAM scores can have unintended consequences that lower quality if teachers are ranked by their VAM scores (ASA, 2014). Variation among teachers' accounts for a small portion of the variation in scores, while most are attributed to other factors such as student background, curriculum, poverty, and other unmeasured variance (ASA, 2014).

In a 2014 joint statement by the American Educational Research Association and the National Academy of Education, those who teach children with disabilities and children that are English language learners tend to have low VAM ratings. Because these children have greater learning challenges than their peers do, the ratings of those who teach them can be low. ASA agrees that test scores are affected by many factors including, (a) the teacher, (b) family, (c) school leadership and resources, (d) class size and curriculum, and (e) student's motivation, attendance, and health (ASA, 2014).

Two changes in the way teachers are evaluated occurred simultaneously. The shift toward VAMs was accompanied by greater differentiation between performance levels. Rather than "satisfactory" or not, performance ratings have three to five levels to differentiate excellence from mediocrity and mediocrity from ineffectiveness (Aldeman & Chuong, 2014). Research suggests a strong association between VAM ratings and principal ratings when using a performance scale with four levels. In addition, appraisals based on VAMs were significantly more predictive of future teacher performance appraisals than performance appraisals based solely on observations.

Value-Added Model vs. Traditional Evaluations

Problems exist when evaluating teacher effectiveness with traditional methods. Infrequent or poor classroom observations or administrator bias affects the validity of teacher evaluations. The subjective nature of traditional evaluations is what feeds the enthusiasm among policymakers for basing teacher evaluation on "objective" test scores.

When Jacob and Lefgren (2008) observed 201 teachers in second through sixth grade, they discovered a strong relationship between principals' evaluations and value-added ratings that were based on student math and reading scores of the same teachers. The researchers then analyzed which method did a better job of predicting how the teachers' future classes would score. They detected that either method was reasonably accurate in predicting which teachers would be in the top and bottom 20% the following year in terms of their students' test scores. Although value-added measures did a slightly better job of predicting future test scores, adding principal ratings increased the accuracy of these predictions. Studies of teacher evaluation systems in Cincinnati, Ohio, and Washoe County, Nevada, also found that value-added measures

and well-done evaluations based on principal observations produced similar results (Milanowski, Kimball, & White, 2004).

There is much complexity and uncertainty in measuring student achievement growth and deciding how much responsibility for gains to attribute to the teacher. To protect teachers from erroneous and harmful judgments, a consensus is emerging among educators that multiple measures are needed that tap evidence of good teaching practices as well as a variety of student outcomes, including but not limited to standardized test score gains. According to a recent study (Coggshall, Ott, & Lasagna, 2010), most teachers support such a multiple-measures approach.

Chapter III: Methodology

Introduction

The purpose of this qualitative case study was to explore teachers' professional and personal perceptions of the VAM to determine teachers' performance, compensation, and personnel consequences. This researcher included the participants' voices, the researchers' instincts and interpretations, and a complex description of the problem (Creswell, 2015). Merriam and Tisdell (2015) stated that education, like other fields such as health and social work, is ideal for qualitative research because it centers on people and everyday problems. Improving conditions is often accomplished by asking questions that can be researched. Merriam and Tisdell (2015) continued, "research focused on discovery, insight, and understanding from the perspectives of those being studied offers the greatest promise of making a difference in people's lives" (p. 1). Furthermore, qualitative researchers want to know how people understand their experiences and how that meaning changes their lives. Merriam and Tisdell (2015) stated, "The overall purposes of qualitative research are to achieve an understanding of how people make sense out of their lives…" (p. 14). This is the primary responsibility of the researcher, who serves as the data collector and interpreter. Good interview questions are those that are open-ended and yield descriptive data, even stories about the phenomenon. The interviewer should avoid multiple-part questions, asking leading questions that make assumptions, or asking questions that only solicit a yes or no response (Savin-Baden & Major, 2013).

VAMs measure student test achievement against the prediction of how students are expected to do given their earlier achievement level and, depending on the specific model, other factors thought to influence student learning that are outside the control of teachers and schools may also be considered in the VAM. Factors such as student poverty or the spending level at a

school are examples of what could be included in the VAM. Research has used the value-added framework to answer questions about the efficacy of various interventions, and the effects of different school resources, such as class size (Hanushek, 1979).

Goldhaber, Walch, and Gabele (2014) explained that VAMs aim to predict what student growth can be expected from an average or typical teacher, and then compare actual student achievement with that prediction. The value-added score for a teacher is intended to illustrate how much the individual teachers contribute to student learning in a specified subject and school year. Teachers who are more productive than the typical teacher are thought to have added value. Teachers are considered less effective when student results show less growth than the typical teacher. VAM measures of teacher performance differ according to the particular VAM used because models differ in terms of how they adjust for student and out of-school factors that influence achievement and the way in which they compare teachers. Some models, for example, predict only student achievement based on prior test scores, while others include factors such as a student's race and ethnicity, eligibility for free or reduced-price lunch, and so on. Teacher performance may compare in correlation to other teachers in the same school or to a larger set of teachers, such as those in a department or even the whole state. The differences between models are sometimes small, but have meaningful impacts on estimates of teacher performance, particularly for teachers who are serving students with backgrounds that differ from those in an average classroom.

According to the FLDOE (2017),

> The teacher's *value-added score* reflects the average amount of learning growth of the teacher's students above or below the expected learning growth of similar students in the state, using the variables accounted for in the model. The teacher's *value-added score* is

> expressed as a sum of two components: one that reflects how much the school's students on average gained above or below similar students in the state (a "school component") and another that reflects how much the teacher's students on average gained above or below similar students within the school (a "teacher component"). FLDOE states,
>
> The most important control, theoretically and empirically, is prior student achievement scores. Students are not randomly sorted into schools or classroom. There are significant differences across schools and classrooms in the entering proficiency of students. A variety of mechanisms contribute to this phenomenon, including parent selection of schools and teachers; teacher selection of schools, subjects, and sections; and principal discretion in assigning certain students to certain teachers. Unbiased estimates of teacher value-added do not require random assignment of students into classrooms (p. 1013).

Evidence that refutes FLDOE's theory that prior test scores take care of unmeasured influences on gains finds that VAM scores tend to be lower according to classroom composition (Haertel, 2013). Teachers' VAM scores do not accurately portray a teacher's effectiveness when their students are more disadvantaged or are low-performing. VAM shows bias against teachers who work with the lowest or highest-performing classes (Haertel, 2013). Many studies state VAM measures appear particularly inaccurate for teachers whose students (a) achieve below or above grade level, (b) are new English learners, and (c) have special needs (Glazerman et al., 2010; Goldhaber & Hansen, 2010; Haertel, 2013; McCaffrey, Sass, Lockwood, & Mihaly, 2009; Newton, Darling-Hammond, Haertel, & Thomas, 2010). VAM is also inaccurate for those teachers who teach in tracked school settings (Harris & Anderson, 2012).

Darling-Hammond (2015) described a situation where two teachers' value-added ratings flip-flopped when they exchanged assignments:

> We had an 8th grade teacher, a very good teacher, the "real science guy" … [but] every year he showed low EVAAS growth. My principal flipped him with the 6th grade science teacher who was getting the highest EVAAS scores on campus, [and] now the 6th grade teacher [is showing] no growth, but the 8th grade teacher who was sent down is getting the biggest bonuses on campus (p. 15).

Darling-Hammond (2015) stressed that constant low ratings occur because certain teachers consistently teach students whose gains are not measured on the grade-level tests; for example, students who are new to the English language, or students in gifted and talented classes (Darling-Hammond, 2015). Teachers account for about 1% to 14% of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system level conditions. Ranking teachers by their VAM scores can have unintended consequences that reduce quality (Chetty et al., 2014).

This qualitative study measured teachers' perceptions of the VAM. Interview questions asked how the faculty perceives the VAM and if they perceive the method as valuable. The questions also addressed teachers' understanding of the VAM and its effect on them personally, and professionally. There is a variety of situations where VAM is difficult to apply. For instance, standardized test scores for students have greater variability from year-to-year due to a variety of uncontrollable factors, including divorce, illness, or similar home-related circumstances. Samuel Meisels (2006) stated:

> Given that young children are undergoing significant changes in their first eight years of life in terms of brain growth, physiology, and emotional regulation, and recognizing that children come into this world with varied inheritance, experience, and opportunities for nurturance, it is not difficult to imagine that a brief snapshot of a child's skills and

abilities taken on a single occasion will be unable to capture the shifts and changes in that development. To draw long-term conclusions from such assessments seems baseless (p. 17).

Meisels (2006) uses two studies to support his claims: LaParo and Pianta (2000) found that only a quarter of variance on academic/cognitive skills on first and second grades tests were accurately predicted by preschool or kindergarten tests. The authors prove that due to the rapidly developing nature of children, standardized testing has little consistent predictive validity. For young children, the authors argue, "instability or change in cognitive and behavioral ability may be the rule rather than the exception during this period" (LaParo & Pianta, 2000). Kim and Suen (2003) performed a similar study and found that "the predictive power of any early assessment from any single study is not generalizable, regardless of design and quality of research" (p. 23) Given the constant developmental changes in a child's brain, both studies illustrate that it is not enough data to form any generalizable correlation for predictability.

Another example is music and art teachers, where no standardized tests exist to measure student progress. If teacher value-added scores cannot be shown as valid for a given purpose, they should not be used. The researcher hopes to shed more light on this aspect of the VAM. A measurement instrument was developed that includes a variety of items about student learning growth, assessment, and the Florida VAM. The FLDOE (2017) adopted the VAM to meet the mandate of the SSA (State Impact, 2014) which required that at least 50% of the annual evaluations of all instructional personnel employed in Florida public schools be derived from student learning growth. The instrument is a semi-structured interview format. Because this study involves human subjects, an approved Institutional Review Board (IRB) application was required. One of the requirements when conducting research with human subjects is to secure

each participant's consent for the study. With the implementation of House Bill 7069 on July 1, 2017, individual districts are allowed to determine how learner growth is measured in teacher evaluation. The district used for this study is utilizing VAM as 35% of a teacher's evaluation for teachers of elementary, English Language Arts, and those that use the Algebra 1 end of course exam. All other classroom teachers for which there are no state assessments, this district is utilizing measurable learning targets approved by individual principals. Prior to the House Bill 7069 implementation, the VAM utilized in a teacher's evaluation was weighted at 50%.

Research Questions

The following research questions serve to focus the literature review, methodology, and further knowledge on the problem statement.

RQ1. How do teachers perceive the use of student's performance on standardized tests in the VAM to determine 35% of their performance appraisal?

RQ2. To what degree do teachers perceive that data generated from the VAM and classroom observations improve teaching practices?

RQ3. To what degree do teachers perceive that performance appraisal using the VAM, weighted at 35%, should inform compensation and personnel decisions?

Methodology

Sampling. To collect data regarding teachers' perceptions about VAM, 12 teachers in a public school district in Florida were interviewed. Stake (2005) suggests that sample sizes for case study research designs include eight to 12 participants, or a sufficient number to reach data saturation. Data saturation occurs when the last interview contributed little new information to address the research question. Adverse selection, either for or against VAM is possible;

however, the nature of the case study design is to reflect the perspectives of the underlying population, in this study, a single school district (Savin-Baden & Major, 2013).

This researcher interviewed a purposive sample of 12 teachers from a single school district: four from the elementary level, four from the middle school level, and four from the high school level. Invitations were emailed to all teachers within the school district to see how many were willing to participate and then twelve participants were randomly selected. The recruitment email included the study title, purpose, confidentiality, and anonymity assurance, and contact information for both the researcher and the IRB. Participants were from 12 different schools to promote data saturation. It was also the desire of this researcher to attain a good sampling of teachers in different professional stages of their careers, annually contracted, and with tenure.

Data collection procedures. A pilot study, or trial run conducted in preparation of a full-scale study, was conducted specifically to pre-test the research instrument to identify potential problems. Various authors have stated the importance of a pilot study as it serves to detect possible flaws in the measurement instrument (Teijlingen & Hundley, 2001; Watson, Atkinson, & Rose, 2007). This is achieved by pre-testing the instrument on a small number of participants having the same characteristics as those in the main study. Sekaran (2003) argues that interviewees can bias the data collected if they do not understand the questions and help to identify unclear or ambiguous statements in the research protocol while Van Wijk and Harrison (2013) believe that pilot studies can add value and credibility to the entire research project. The three participants in the pilot study provided validity on whether the instrument needed an adjustment.

The researcher used the same criteria for the selection of participants for the pilot study as the main study. The target participants in this pilot study were teachers that taught VAM-

tested subjects at the high school and middle school level. When using an interview as a research tool, particularly face-to-face interviews, it is best to utilize a setting that provides the most comfort for the participant (Jacob & Ferguson, 2012). However, the setting must also be without too much background noise or distractions so that recording of data is made easier. They also suggest setting aside uninterrupted time that is estimated to be adequate to complete the interview. For this reason, participants are allowed to choose the setting for the interview that they are most comfortable with so long as they would also provide a quiet environment. The majority of participants have chosen their classroom or the teacher's lounge during their lunch or teacher planning time.

When conducting any research, not only is the choice of an appropriate data collection instrument very important, but more important is to ensure that the chosen instrument performs the desired job properly (i.e., collects the right data). This is even more paramount in qualitative research where, unlike with quantitative, data is neither exact nor statistical in nature and therefore requires ensuring that instruments capture required concepts. Participants for the pilot study, though reflective of the target respondents in the main study, were chosen purposively based on what they teach, the convenience of access, and willingness to participate in the pilot.

Study data was collected using in-depth, semi-structured interviews and journaling. Twelve Florida public school teachers were recruited to participate in this study. Case study research designs involve small sample sizes accompanied with in-depth interviews to reach data saturation. Data saturation occurs when the addition of one participant provides little marginal information (Savin-Baden & Major, 2013). Interviews were one-on-one and at a mutually convenient location (Stake, 2005). Participants were notified that interviews would be recorded. Semi-structured interviews lasted approximately 15 to 30 minutes and included questions in

Appendix A. Interview questions were structured to enable in-depth responses from the participants that communicated experiences and perceptions in their own words (Lincoln & Guba, 1985). A journal was maintained by the researcher to collect non-verbal data from interviews with participants, and to reflect personal observations.

Data analysis. Semi-structured interviews were transcribed into MS Word format and emailed to participants for review and editing to ensure accuracy in a process referred to as *member checking* (Savin-Baden & Major, 2013). The first pass through the transcribed interviews involved recording and summarizing participants' demographic data on a spreadsheet for age, gender, ethnicity, highest education level, teaching experience, and present grade level. Interviews were analyzed using content analysis to identify and code themes, patterns, ideas, and phrases (Glaser & Strauss, 1967; Lincoln & Guba, 1985). A code is a word or phrase that summarizes the portion of recorded data; the goal of the researcher is to understand the repetition of the codes to make meaning of the patterns and understand the links among the data. The act of coding is not "a precise science" but rather an "interpretive act." Coding is exploratory and requires one to discover the patterns and features of the data, categorize them, and arrange them systematically and efficiently (Merriam & Tisdell, 2015).

After several readings of all participant interviews, blocks of verbiage were copied into a spreadsheet and assigned meaningful headings and subheadings derived from the content based on a common element. Headings and sub-headings were derived based on common themes, ideas, or phrases. The headings served as categories to organize and analyze emerging themes and reflect recurring language, feelings, perceptions, or experiences. As the content analysis proceeds, headings, and subheadings evolved to reflect the accumulated content, and contiguous patterns and themes were combined (Savin-Baden & Major, 2013). The resulting matrix of

participants against themes and sub-themes reflects participants' experiences and worldview in a manner that promotes insights into the underlying phenomena that would not otherwise have been possible. Matrixes summarize relationships between individuals and clusters with common demographics, themes, or experiences. The aim of this study was to find at least four to five themes that support the research questions and reflect participants' interviews.

The researcher used the constant comparison method to organize and analyze the data. The method of comparing and contrasting is used for forming categories, establishing boundaries for each category, assigning words and phrases to categories, summarizing the content of each category, and revising categories in an iterative process until coherent themes emerge (Glaser & Strauss, 1967). The researcher reviewed and coded printed interview transcripts and used the constant comparison method of content analysis to collect and organize the data. Descriptive statistics (means, frequencies, standard deviation) were used to characterize the study sample and identify overarching patterns and commonalities among participants. Finally, categories and subcategories underwent content and definition changes as units and incidents accumulated, and as category properties changed or relationships between categories developed during the simultaneous data collection and analytical processes.

Role of the Researcher

The role of the researcher is critical in qualitative research because the researcher is the main instrument for collecting data, serving as both observer and investigator (Yin, 2013). Consequently, the researcher must explicitly understand the potential for researcher bias. Each person brings the catalogue of their experiences to each new interaction and there is potential for projection or transference to occur in data collection and analysis. Journaling was employed to minimize the potential for researcher bias. A journal was used to record personal thoughts and

assumptions during the data collection and analysis processes. Journaling helps to mitigate against preconceptions and biases that may taint study findings (Tufford & Newman, 2010). Gearing (2004) described journaling as a "scientific process in which a researcher suspends or holds in abeyance his or her presuppositions, biases, assumptions, theories, or previous experiences to see and describe the phenomenon" (p. 1430). The use of journaling enabled the researcher to make conscious preconceptions to reduce the potential for bias.

Summary

The purpose of this qualitative case study was to explore teachers' professional and personal perceptions of the VAM to determine teachers' performance, compensation, and personnel consequences. Data was collected from a purposive sample of 12 teachers from a single Florida school district using semi-structured interviews. The researcher used journaling, and the constant comparison method to minimize the potential for researcher bias. Content analysis was used to analyze interview data and address research questions. Content analysis involves the derivation of common themes, ideas, or phrases through an iterative process of re-reading and coding. The study sample was characterized using descriptive statistics and study findings were organized by research question. Chapter IV contains information regarding the study sample and findings by research questions. Chapter V concludes with the discussion of the findings and what the findings reveal about the participants as well as what it means in relation to the theoretical framework. This chapter will also address the limitations of the study and recommendations for change.

Chapter IV: Results

Introduction

The purpose of this qualitative study was to explore both teachers' professional and personal perceptions of the value-added model (VAM) to better understand teachers' performance, compensation, and personnel consequences. VAM is used to evaluate teachers' performance and includes improvement in students' performance on standardized exams as an important criterion (Guarino, Reckase, & Wooldridge, 2015). The use of VAM to evaluate teachers' performance was replaced by previous evaluation models as indicated in the 2011 statute (FLDOE, 2017). However, there have been concerns introduced regarding using VAM as a means of judging teacher performance (Amrein-Beardsley, Pivovarova, & Geiger, 2016). Chapter IV includes a discussion of the data collection methods used as well as a discussion of the findings. Descriptive and demographic statistics for the participants are included, and study findings are summarized. In Chapter V, conclusions and recommendations will be introduced.

Data Collection

Data were collected using in-depth, semi-structured interviews and journaling with a purposive sample of 15 teachers recruited from a single public school district in Florida. First, a pilot study was conducted in preparation of the full scale study to specifically pre-test the research instrument to identify potential problems. Participants for the pilot study, though reflective of the target respondents in the main study, were chosen purposively based on what they teach, the convenience of access, and willingness to participate in the pilot. The overall result of the pilot test was that the interview protocol satisfied the requirements for validity as it could adequately be used to obtain data on the concepts that this researcher hoped to measure. Therefore, the result from the three teachers of the pilot study were combined with the results of

the actual research. The 15 teachers utilized for this study consisted of two elementary teachers, six middle school teachers, and seven high school teachers. Since only two elementary teachers chose to participate, one middle school and one high school teacher that do not utilize VAM tested subjects were chosen to obtain their perceptions. Sample demographic information is summarized in Table 1.

Semi-structured interviews were one-on-one, lasted between 15 and 30 minutes, and permitted respondents to explore their own perceptions. The qualitative data collected from the interviews was then analyzed using content analysis and themes identified that emerged in the responses. The creation of themes following content analysis involved several readings to ensure that consistent themes were identified. All interviews were conducted in person, recorded, and transcribed and member checked to ensure accuracy. Participants were assigned pseudonyms and any personally identifiable information is stored separately on a password protected offline storage media. Study data were stored in password protected file on a password protected removable storage device in a locked drawer.

Data Analysis

Interviews were transcribed into MS Word format and emailed to participants to ensure accuracy in a process referred to as *member checking* (Savin-Baden & Major, 2013). The transcription process included multiple passes to ensure that accurate information was collected. Data acquired included demographic data as well as interview data that was used as the basis for content analysis. Coding for themes occurred as the researcher watched for the repetition of certain phrases in order to ascertain patterns in responses. As similar phrases were identified, these were linked together where appropriate. This process required that the researcher be observant to identify clusters of similar phrases and how participants connected ideas. Several

readings of the interviews occurred and blocks of verbiage were transferred into a spreadsheet. This verbiage was later assigned headings and subheadings that appropriately summarized the clustered verbiage, with verbiage under these headings and subheadings appropriately relevant to those headings as determined by a reading of their content. The creation of these headings was an iterative process that led to new headings emerging with subsequent re-readings, and the final headings thematically summarized the experiences of the participants in this study. A comparison method was used during this process, which involved coding transcripts and using a constant comparison to organize the collected data.

Sample Demographics

Two elementary school teachers, six middle school teachers and seven high school teachers participated in the study sample. As shown in Table 1, study sample included three (20%) males and 12 (80%) females ranging in age from 27 to 61 with a mean age of 49.2 (*SD*=8.2) years. The sample mean annual salary was $50,320 (*SD*=$7,188). The sample educational attainment level included five (33%) masters' degrees, eight (53%) bachelor degrees, and two (13%) specialist degrees. The study sample included 10 (67%) tenured teachers and five (33%) with annually renewable contracts. The study sample mean teaching experience was 17.3 (*SD*=6.0) years. The study sample included 10 (67%) White, three (19%) African Americans, one (7%) Hispanic, and one (7%) Asian.

Table 1

Study Sample Demographics

	Subject	Teaching Experience	Degree	Ethnicity	Age	Gender	Salary	Contract Type
1	8th ELA	12	Specialist	AA	55	F	$47,000	Tenure
2	9 & 10th ELA	15	BA	Asian	47	F	47,000	Annual
3	9 thru 12th ELA	17	Specialist	White	49	F	50,000	Tenure
4	9th Algebra	14	BA	White	36	F	45,000	Tenure
5	9th Algebra	20	BA	White	61	F	65,000	Tenure
6	4th ELA	28	MA	White	52	F	50,000	Annual
7	8th Ph. Science	23	MA	White	56	M	51,000	Tenure
8	8th ELA	20	MA	White	44	F	49,800	Tenure
9	9 thru 12th Arts	18	BA	White	54	M	47,000	Annual
10	6 thru 8th Math	24	BA	White	52	F	48,000	Tenure
11	8th Math	20	BA	White	48	M	49,000	Tenure
12	4th Math/Science	5	BA	Hispanic	27	F	41,500	Annual
13	10th ELA	6	BA	AA	54	F	41,500	Annual
14	9 thru 12th Reading	20	MA	White	55	F	68,000	Tenure
15	7 & 8th Reading	18	MA	AA	48	F	55,000	Tenure
Mean		17.3			49.2		$50,320	
SD		6.0			8.2		$7,188	

N=15

Results

RQ1. Do teachers believe the use of student standardized test performance should be used in the VAM to determine 35% of their performance appraisal?

The first theme that emerged from the semi-structured interviews was the poor communication efforts by administrators regarding VAM. Teachers were unaware that the percentage of the VAM used in their evaluation changed from 50% to 35%. Respondent Number 8 said, "35%? You see, that right there. The evaluations are always changing. This is how pathetic and how well everything is explained to us." Respondent 1 also felt

communication from administration was poor when their response was, “What is that 35%? What of that 35% represents my performance? I just don't know.” Even though 35% is lower than the original 50%, the majority of teachers felt it accounted for too much of their performance. As Respondent 2 said,

> The VAM is 35%? I don't think that it's fair because a teacher can teach all the goals and standards, but what if a student doesn't care…I can give the best lesson, but 35% of the VAM....I just think they need a different evaluation system.

There was a generally negative reaction to the VAM accounting for such a large portion of their performance evaluations. In some cases, this was simply because using a single state exam to account for such a large percentage was viewed as unfair. In other cases, teachers felt using the VAM to account for so much of their performance was poorly thought out since they did not understand how they were scored. As Respondent 3 indicated, “It's not fair. Until I can completely understand and verify the math myself, it is an arbitrary tool put in place that does not allow for either verification or deeper understanding.” How VAM was calculated was left nebulous rather than communicated effectively. How VAM should be applied also differed from one administrator to another. While teachers were able to grasp the general idea of how VAM was applied, many lacked concrete ideas of the process and its outcomes. With regard to the uneven application of VAM, Respondent 2 commented,

> One assistant principal can see something totally different from the principal. One can get innovative, one can get applying. There is no one training us or even having a video saying, “This is what this one learning or standard should look like.”

Respondent 2 indicated that a lack of communication at a leadership level meant that different leaders had different views of VAM and its application. Worse, there was a lack of a

cohesive message that filtered downward to the staff. This left its integration lacking. As said by Respondent 3, "VAM remains a mystery; therefore, it really has not impacted my teaching. Rather, self-studies, strategies, and professional learning communities have had greater impact." The lack of communication regarding VAM left the understanding of it beyond the school staff, and they were unable to use it as a means of addressing their instructional methods. Teachers also felt that it was unfairly leveraged against them. As Respondent 8 said,

> There is no explanation for how it's calculated at all. A lot of teachers using VAM may have certain groups of students that are not capable of reaching those high levels. So we are evaluated on a mysterious formula using methods we are told to use and it is very easy to get punished for that.

Respondents felt that communication was lacking regarding VAM on several fronts. Comments indicated that VAM was communicated in a broad and inexact way regarding how it was calculated, applied to them, and how to use it to address their teaching. Some teachers commented that they felt there should be more involvement from above the administrative level, at the county level, to better train and educate regarding VAM.

RQ2. To what degree do teachers perceive that data generated from the VAM and classroom observations improve teaching practices?

A specific way that data generated from VAM confused respondents is in the way that the findings are presented. Most respondents felt the VAM reports were indecipherable, inaccurate, or unrelated to classroom teaching practices. A sample of a final VAM report, with all identifiable information removed, is included in the appendix. Some respondents felt that VAM scores over-weighted students' standardized test scores relative to other means of gauging teacher performance, such as pass rate. Beyond respondents' confusion regarding VAM scores,

no respondent understood how the score informed instructional practices. On point, Respondent 4 said, "I know when I get my scores and I can see that they are good and the VAM comes out as a positive number, but it doesn't inform me on a daily basis about my instructional practice."

Teachers remained unable to integrate the scores they received into their instructional methods. Although VAM should have been able to show them areas of weakness, they lacked an understanding of how to interpret and integrate the test scores. In some cases, teachers doubted whether it appropriately gauged their work at all. As stated by Respondent 5,

> Definitely, test scores are a huge factor which I don't necessarily think are fair…I had a principal where I had 100% pass rate and this principal called me in and said that was not a very good score and you are just an average teacher. So he scored me average and at that point, I quit caring about the VAM score. I simply only cared about my students so that they could pass and move on.

What should have been a tool by which teachers can improve instead remained a mysterious metric that they felt was used against them inaccurately. This led to them abandoning confidence in the process. The unfair nature of how VAM was used was not only attested to by Respondent 5. Respondent 11 said,

> I don't think it's fair and accurate. Before VAM, I was always rated a 4 (highly effective) from the state. Since the VAM has changed to what it is now, I haven't gotten above a 3.25 (effective) from the state. Once they changed it, it has always lowered my scores. Not only that, nobody can explain it clearly on how we are judged.

Teachers resented VAM not just because they did not understand how to use it as a means of informing their instructional methods, but because it was not consistent with other means of assessment. Whether compared to other existing metrics or to past methods of

assessment, the VAM seemed to suggest that teachers were not as effective as those previous evaluation metrics. The disparity created anger toward the method of scoring VAM, particularly since there was little explained to teachers about the method of calculation. The VAM was perceived not only as confusing and unfair, but respondents had no knowledge regarding how best to apply VAM scores to improve their teaching methods. Respondents also disliked the disparity between VAM and other assessment metrics that rated them more positively.

While the majority of respondents felt that the VAM had not helped improve their performance, there were some positive responses to the VAM. Some teachers indicated that it had helped them identify areas of weakness and ways to do their job better. These individuals indicated that the VAM had helped them find yearly goals to work toward as well as shortcomings they could improve upon. This helped some target students' weak areas for improvement. As Respondent 2 said,

> The VAM showed me student growth in reading gains. The students that did not pass, I could see what their weak areas were based on the FSA scores. It shows areas of the curriculum I could have worked on more.

Respondent 2 felt the VAM helped to identify areas of weakness within the student population. While some respondents felt that the VAM didn't help to inform their instructional improvements, those who did feel they benefited from the VAM indicated that it helped them to find weaknesses among their students. As mentioned by Respondent 7, "It gives me some things to work on that you want to choose as a yearly goal for improvement." Teachers were able to identify ways to improve in at least yearly time spans. However, the ability to improve on the basis of the VAM was limited given the lack of understanding regarding the VAM score. Respondent 13's response hinted at this lack of understanding when they said, "The parts I do

understand are the things that go on in the classroom. It helps me to stay focused and make the modifications I need."

Eight of the respondents indicated that the VAM had not helped them to improve. The lack of communication and inability to appropriately interpret the VAM seemed to manifest in a general belief that the VAM did not help teachers perform better. For many, it seemed to hold no value as a tool that could be used to adjust instructional performance. However, a few members did indicate that they were able to interpret the VAM to the degree that they were able to identify weaknesses in their instructional methods and make adjustments. They were able to identify weaknesses in the instructions they delivered to student and establish ways to improve.

RQ3. To what degree do teachers perceive that performance appraisal using the VAM, weighted at 35%, should inform compensation and personnel decisions?

The strongest theme to emerge from this research is that respondents felt the Florida VAM teacher evaluation system was unfair. Some noted that student achievement was a multi-faceted outcome that involved far more than teachers' input. Others felt that they couldn't see a connection between merit pay and VAM, since they didn't understand the VAM in the first place and did not understand how it was calculated. Others felt that merit pay pitted teachers against one another in a competition. The system fostered a negative environment in which some teachers resented others who had the fortune of receiving less challenged students that year. Speaking to the fact that merit pay was often unfair because of circumstances beyond a teacher's control, Respondent 6 said,

> When you compare a classroom full of Exceptional Student Education students (ESE) and English Language Learners (LY) compared to the class across the hall that is full of

gifted kids that are going to score well pretty much regardless of the teaching they receive. That's not fair to me.

In this instance, merit pay seemed unfair because the circumstances and contexts in which teachers instructed could be very different. A more difficult class might make performance more difficult. Consequently, it was harder to earn merit pay when circumstances were made more difficult based on class composition. Adding to this, Respondent 10 added that personal relationships may impact assessments and, thus, merit pay. Respondent 10 said,

> Some of the evaluations from some of the administrators can be subjective. Unfortunately, if you have a great relationship with a certain administrator, then you may have a higher evaluation, if you don't get along with them so well, you may have a lower evaluation....it's too subjective so I don't think it's fair at all. It doesn't motivate me to improve...it does the opposite for me.

In this second instance, merit pay was perceived as unfair because teachers felt that personal relationships could impact the outcomes of evaluations and influence merit pay. Teachers therefore felt that merit pay was unfair not only because class composition impacted evaluations, but also because relationships between administrators and teachers affected relationships. Finally, if merit pay was supposed to motivate teachers, it did not seem to do so. As Respondent 11 said,

> It's not fair. Once they went to merit pay, everything becomes personal. When they tie evaluations to money, it becomes personal. Putting merit pay based on a score that they cannot explain to us and is not used on every teacher, it becomes an unfair system. It doesn't motivate me.

Merit pay was sometimes perceived as unfair, while at other times teachers didn't see the connection between merit pay and VAM at all. Teachers felt that student achievement was something they could only control to a degree. When the odds were stacked against them, such as when their classes were full of students who were more highly challenged, it made the concept of merit pay seem unfavorable. They also felt that the influence of personal relationships on merit pay made the system unfair. As a means of motivating these teachers, the merit pay method seemed to fail in its goals.

Florida Statute SB736, that passed in 2011 in Florida, requires all newly hired teachers receive one-year, renewable contracts, thereby requiring teachers to be rehired annually with tenure eliminated. The notion emerged that the current teacher shortage in Florida schools could be directly related to using the VAM for personnel consequences. In regard to job retention, many teachers did not feel that this was something that applied to them as ten out of the 15 teachers interviewed have tenure. For example, Respondent 9 said, "I have no idea how they are using it. I think it can be manipulated to fire who they want." Respondent 1 concurred with their response by stating, "I have no idea in terms of job retention. I didn't even know it was related to job retention. I have no knowledge about that." Respondent 4 indicated, "I would think that because I have tenure that they are not going to fire me because I happen to have a bad year with some low score. But, I imagine if it was a repetitive thing for multiple years, they would probably not want to hire me back. Is that fair? I don't know about all of that."

A few teachers brought up the concern that using the VAM for job retention would push out good or new teachers. Respondent 7 said, "I have a colleague that's not tenured and she freaks out every time these things come out, she's afraid she will lose her job." As Respondent 10 put it,

> Unbeknownst to the district, they are pushing away a lot of good teachers. There are a lot of new teachers coming out of college that want to make a difference… and they find out all of this legality… it crushes new teachers.

The use of a single test was perceived as a means of pushing away teachers. Rather than being a means of improving teaching performance and encouraging retention, some teachers felt it had the potential to drive staff away.

In summary, the strongest theme is that teachers felt that the Florida VAM evaluation system was unfair. Student achievement is a multi-faceted outcome that involves factors far beyond the control of the teacher. Teachers could not see a connection between VAM score and their teaching practices and most did not understand the VAM calculation. Overall, teachers perceived that the VAM system fostered a negative environment in which some teachers resented others who had the good fortune to receive fewer challenged students in any given year.

Potential Biases

There are potential biases that could impede the data for this case study. This investigator's teaching and educational experiences could play a role. This researcher is still a teacher and maintains contact with current teachers to discuss educational issues. Another potential bias is that this researcher has previous and current experience with standardized testing and teacher evaluations that utilize VAM. In addition, the current research from the literature review in this case study in regard to the validity of VAM has shaped this researcher's opinion and view of the VAM. By being aware of these potential biases, this researcher utilized a reflective journal during data collection to avoid bringing personal biases into the interviews. No comments or facial expressions were made by the investigator during the collection of data.

Limitations

The first limitation of this study is the use of only one school district in Florida. Since only one school district was used, the findings only represent the perceptions of 15 teachers in this particular school district and generalized findings outside of this population in this study may not be possible. The second limitation is that it is possible that the teachers' responses in regard to the demographic question on salary does not accurately reflect their true salary. When asked about salary, there were long pauses and responses that indicated that they were not entirely sure. To clarify, four teachers did respond that they were not sure as they do not teach for the money or they were not sure of their base salary due to extra jobs. Upon emailing the teachers after the interview to verify the salary, only one teacher replied back with a different amount. According to the salaries reported by the teachers, the data reveals that there are inequalities in how salaries are distributed across teachers, but merit pay may play a role.

Chapter V: Discussion, Implications, and Recommendations

The purpose of this qualitative study was to better understand teachers' perceptions regarding value-added modeling (VAM) and better understand consequences to personnel, performance, and compensation. The VAM was implemented in Florida following passage of Senate Bill 736 in 2011 (Florida Statute, 2011), and this research provides an exploration of teachers' perceptions regarding the VAM used in their personal evaluations. This study occurred only a few years after its passing, making it timely with regard to adding to the current understanding of the VAM's impact. School level data was collected via the use of semi-structured interviews conducted among 15 public school teachers from a school district in Florida. The three teachers for the pilot study, though reflective of the target respondents in the main study, were chosen purposively based on what they teach, the convenience of access, and willingness to participate in the pilot. The overall result of the pilot test was the determination that the interview protocol does satisfy the requirements for validity as it could adequately be used to obtain data on the concepts that the researcher hoped to measure. Therefore, the data from the pilot study was compiled with the data of the twelve teachers that participated in the study. The 15 teachers consisted of two elementary teachers, six middle school teachers, and seven high school teachers. Since only two elementary teachers chose to participate, one middle school and one high school teacher that do not utilize VAM tested subjects were chosen to obtain their perceptions. In collecting the data, journaling was also utilized.

Discussion

Qualitative data were generated through the three overarching research questions by utilizing several steps. First, an invitation was sent out to administrators from a school district to be forwarded to their teachers inviting them to participate in the study. Fifteen participants

volunteered to be interviewed. The data collected from the interviews were transcribed, coded, and analyzed using the traditional method of identifying patterns and themes through data analysis. The findings were then written and presented in Chapter IV. Chapter V provides a general discussion, implications of the findings, recommendations for policy, practice, and further research.

Overall, the findings from the data are not favorable towards VAM being used in teacher evaluations. Data from the semi-structured interview questions revealed that the majority of teachers have limited understanding of the VAM. Teachers are aware that their scores are from their student's test scores on state exams, but they are not sure how their VAM score is calculated and what variables are included in the VAM. Other negative feelings towards VAM derived from teachers feeling that they had no control over variables such as a large number of students with learning disabilities, home life problems, lack of motivation, etc. The lack of control over such variables also leads to teachers feeling that merit pay in connection to their VAM scores is unfair. Teachers were also unaware how VAM scores impacted personnel consequences as it has not impacted them. The strongest attribute of VAM being used in evaluations is that VAM scores allow teachers to understand student weaknesses so that they could target their instruction toward improving those weaknesses.

The following research questions were formulated to guide this research:

RQ1. Do teachers believe the use of student standardized test performance should be used in the VAM to determine 35% of their performance appraisal?

For Research Question One, open-ended questions related to teachers' perceptions on whether they believe VAM should be used to determine 35% of their performance appraisal, were utilized. The responses were closely aligned to their experiences as instructional leaders as

well as the culture of their school district. All 15 teachers interviewed, recognized that teacher evaluations are utilized to hold educators accountable for the instructional choices they make in their classrooms. Many of the teachers interviewed were unaware that their school district changed the percentage of their VAM evaluation from 50% to 35%. The lack of communication from district and administration regarding VAM left a misunderstanding of it, and they were unable to use it as a means of addressing their instructional methods. As Respondent 3 indicated, "It's not fair. Until I can completely understand and verify the math myself, it is an arbitrary tool put in place that does not allow for either verification or deeper understanding." As a result of the lack of communication, teachers did not understand how VAM was calculated, how to integrate it in such a way as to improve their instructional methods, or how it was used to assess their performance. When communication was made, different leaders communicated different things, confusing teachers as to what they should know. Poor communication therefore had a negative systemic impact on how teachers perceived VAM and its impact on their performance. Teachers also felt that 35% was unfairly leveraged against them. Given that so much of their performance was gauged by this single metric, it left them feeling as if the rest of their work was being undervalued with regard to student improvement.

RQ2. To what degree do teachers perceive that data generated from the VAM and classroom observations improve teaching practices?

In response to the open-ended questions that focused on Research Question Two, some of the teachers indicated that their VAM score helped to highlight areas of weakness in their students' growth and helped them to more closely target these weaknesses. As such, the VAM was perceived as helping teachers to more specifically address students' needs in order to maximize the degree of improvements within a class. Respondent 4 said, "I know when I get my

scores and I can see that they are good and the VAM comes out as a positive number, but it doesn't inform me on a daily basis about my instructional practice."

The scores also set yearly targets for teachers. By better understanding areas of weakness from the previous year, teachers were able to improve their instructional methods so that they were able to become better instructors and shore up their own instructional weaknesses. Conversely, the majority of the teachers were simply confused by the VAM and provided no extensive answers to questions regarding how VAM could help them improve. As such, nonverbal cues suggested these teachers were dismissive of VAM and how it could help them improve, again, rooted in the fact that they did not understand their scores.

RQ3. To what degree do teachers perceive that performance appraisal using the VAM, weighted at 35%, should inform compensation and personnel decisions?

Class size and classes with more students who had learning difficulties were likely to produce lower VAM scores. As a result, teachers felt VAM was an incomplete and unjust method of assessing their performances because their perception is that teachers with smaller classes and fewer students with learning disabilities would typically be scored higher which ultimately means that those teachers with students with the most learning disabilities are less likely to receive merit pay.

Respondent 6 said,

> When you compare a classroom full of exceptional student education students and English Language Learners compared to the class across the hall that is full of gifted kids that are going to score well pretty much regardless of the teaching they receive. That's not fair to me.

Haertel (2013) emphasizes that students should not only vary by general cognitive abilities, but should also vary in relevant prior experiences that may also impact achievement of students. Evidence from Conley and Glasman (2008) suggests teachers only feel comfortable being evaluated on variables within their control within their classroom.

Implications

The use of a qualitative study complicates generalizability in a study. Qualitative studies such as the current one draw on a small sample size, limiting the ability to generalize the experiences of participants to the larger population. The experiences are too limited in nature to be able to apply across the experiences of a larger population. The ability to expand the findings to a larger population is also restricted given the geographic concentration of the teachers within a single school district. The conditions specific to a single school district may not apply across a wider number of districts. There may be unique variables specific to the district not encountered in other districts. As such, the findings of the research are also geographically limited and may not be applicable for teachers across a wider geographic area because this study reflects the perceptions of 15 teachers in one, single school district in Florida. It is not unreasonable to consider the possibility that teachers from other schools and districts may have similar perceptions as those reflected in this study. In order to validate the possibility, further research on a much larger scale would need to be conducted.

Also, the study was limited by the nature of the data collection. The semi-structured interview protocol required teachers to respond to several related questions regarding the VAM. However, the interview format of a study can influence the responses that a participant provides. Participants may provide answers and responses that they believe are expected of them by the researcher. As part of complying with ethical concerns, teachers were informed of the nature of

the study and also assured regarding the anonymity of their identities. However, the point in assuring teachers of their anonymity was partly to encourage honest responses to the questions. Teachers were assured that there would be no negative consequences for their responses, therefore the researcher made the assumption that participants responded honestly.

Recommendations for Practitioners

With regard to the immediate data collected in the current study, school administrators could apply the findings in such a way that it informed a more effective implementation of VAM within their schools. Given that one of the recurring findings was a lack of understanding regarding VAM calculations and how they could be used to alter instructional methods, one of the first steps in improving VAM application may be by better informing teachers about what VAM scores indicated and how they could improve instruction. By doing so, it might make the entire process of integrating VAM into a school evaluation more widely accepted and effective with regard to improving instructional approaches. With the implementation of House Bill 7069, which now allows districts to determine how learner growth is measured (Florida Senate, 2018), legislators recommended that teachers, principals, curriculum specialists, union representatives, students, and parents be part of a committee designing the performance appraisal. To promote teachers' understanding of VAM calculations, administrators could publish clearly defined variables used to calculate VAM scores; develop a VAM training curriculum and educate teachers' and administrators together to promote the same interpretation of the VAM evaluations; and provide vignettes or examples of how the VAM calculations could be used to improve instructional approaches.

To provide an accurate picture of a teacher's performance in the classroom, it is recommended that multiple measures be utilized in a teacher's performance appraisal (Goe, Bell,

& Little, 2008). Allowing multiple measures in a teacher's evaluation will allow school districts to make fair comparisons among teachers when determining merit pay. As teachers indicated in this research, merit pay was unfair when solely utilizing the VAM from a single state test as a determination for additional pay. Teachers also indicated that being measured by a single assessment made them feel that all of their work to attain student achievement was undervalued. Additionally, since schools cannot generate VAM for all teachers, it results in a lack of achievement data which is also unfair for performance-based compensation. If multiple measures are utilized to evaluate teachers on a wide range of assessments, it would enable districts to identify highly-effective teachers for compensation decisions.

Finally, sophisticated formulas exist that can adjust VAM scores to adjust for disparities in the proportion of SPED or ESL students. Such adjustments to an evaluation system based on student performance are necessary due to inequitable distribution of low performing students to teachers. The potential benefit of using sophisticated formulas to adjust VAM scores for disproportionate distributions may be diminished by the incremental complexity of adding a "black-box" adjustment to VAM scores. The highest and best use of VAM evaluations is to improve student performance by measuring and improving teacher's mastery, and simplicity and a connection to the classroom is necessary.

Recommendations for Academics

The current study was based in qualitative research involving the creation of various categories and themes as derived from semi-structured interviews. However, no attempts were made to determine correlations between various variables. Future research could expand to a quantitative format to determine correlations between categories, such as linking merit pay to student academic outcomes. Other changes to the format could also include attempts to increase

the generalizability of the study. This would include attempts to broaden the geographic scope of the study, include a more diverse number of districts, and widen the number of participants. By doing so, particularly in a quantitative study, it would be easier to arrive at conclusions that could be generalized across the wider teaching population. Perceptions regarding VAM would be easier to generalize across larger populations through this expansion of the study's scope.

Additional research could be done by comparing other states to Florida that also utilize VAM. This could bring further credibility to the findings or refute this research if contrary findings emerge. In addition, another area could be to compare Florida's model to another state that utilizes a vastly different evaluation system other than VAM, utilizing the same research questions but omitting VAM in the question and inputting their model. By doing this, it could possibly find a more effective evaluation system to implement.

Since the new Florida House Bill 7069 states that districts are allowed to determine how learner growth is measured (Florida Senate, 2018), the VAM formula is now open to interpretation and at the discretion of local school boards and districts. Future researchers could examine how different Florida school districts are determining what will be used in a teacher's evaluation and what the effect would be on a teacher's performance appraisal.

Conclusions

There were several key findings arrived at following analysis and organization of the qualitative data. Teachers' perceptions of the shortcomings of VAM evaluations included: a) VAM calculation was poorly communicated, resulting in confusion regarding how to maximize performance; b) inequitable distribution of low performing students creates bias in VAM scores; c) a single measure of academic improvement failed to capture important dimensions of learning; d) VAM scores were unconnected with instructional improvements. VAM has the potential to

inform teacher instructional methods in order to produce improvements. However, proper use of VAM is required to effectively integrate VAM into their teaching. While there are significant limitations to this study, the current findings indicate that there are numerous ways in which the VAM might be integrated. Poor communication seemed to underpin a number of different problems that teachers had with VAM, including using it to improve their instruction. Administrators can use these findings to inform the integration of VAM into their schools. Properly communicating is important to helping teachers improve their instructional methods.

By clearly communicating how to interpret scores, administrators can help teachers become more responsive to the VAM and use those scores to improve their teaching. However, administrators also need to concern themselves with issues of fairness, since many teachers felt that class composition and size, both factors that impacted their performance, went unaccounted for in VAM. The current study reveals that issues of communication must be addressed when VAM is integrated and that teacher concerns surrounding fairness must be taken into account. While the VAM can help to improve performance, teachers' various issues must also be addressed in the process.

The goal of the teacher evaluation process is to ensure high quality education inside the classrooms to drive student achievement increases (Weems & Rogers, 2010). Teaching is a complex job that has serious implications and the future of our nation needs highly effective teachers. Ineffective teachers could cripple our nation, so accountability in the classroom is imperative. Gallagher (2004) states, "As schools and districts across the country work to improve student achievement, it is important that high quality teaching and high-quality teachers be identified." Many forms of evaluation could be used to determine teacher quality such as principal observations, peer evaluations, student scores on various tests, portfolios, surveys,

student/parent feedback to name a few. To protect teachers from harmful judgments, a consensus is emerging among educators that multiple measures are needed that provide evidence of good teaching practices as well as a variety of student outcomes, including but not limited to standardized test score gains. According to Coggshall et al. (2010), most teachers support such a multiple-measures approach. Even though errors can occur in all of these sources of data, it is vital that validity evidence is collected to ensure the quality of teachers in our classrooms.

This research looked at one aspect of how complex the process of teachers' evaluations are through the utilization of the VAM in teacher evaluations. Any measurement in regard to teacher evaluations will have its challenges. However, this study has brought about some important points for consideration with regard to teacher evaluations that will need ongoing research with the need of involvement from teachers, students, parents, politicians, and various stakeholders.

References

Aldeman, C., & Chuong, C. (2014). Teacher evaluations in an era of rapid change: From unsatisfactory to needs improvement. *Bellwether Education Partners*, 1-40. Retrieved from https://eric.ed.gov/?id=ED553852

American Institutes for Research. (2013). Florida value-added model technical report. *Working Draft for Review and Comment,* 1-80. Retrieved from https://www.ocps.net/UserFiles/Servers/Server_54619/File/Departments/Research, Accountability and Grants/Accountability, Research and Evaluation/Value-Added-Model-Technical-Report-AIR.pdf

American Statistical Association. (2014). *ASA Statement on using value-added models for educational assessment.* Retrieved from http://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. New York, NY: Routledge.

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS® EVAAS®) in the Houston independent school district (HISD): Intended and unintended consequences. *Education Policy Analysis Archives, 20*(12), 1-23. Retrieved from http://www.redalyc.org/html/2750/275022797012/

Amrein-Beardsley, A., Pivovarova, M., & Geiger, T. J. (2016). Value-added models: What the experts say. *Phi Delta Kappan*, *98*(2), 35-40. doi.org/10.1177/0031721716671904

Barton, S. (2010). *Principals' perceptions of teacher evaluation practices in an urban school district* (Doctoral dissertation). Available from ProQuest Dissertations and Theses.

Booher-Jennings, J. (2005). Below the bubble: Educational triage and the Texas accountability system. *American Educational Research Journal, 42*(2), 231-268. doi:10.3102/00028312042002231

Bouchamma, Y. (2005) Evaluating teaching personnel: Which model of supervision do Canadian teachers prefer? *Journal of Personnel Evaluation in Education, 18*(4), 289-308. doi:10.1007/s11092-007-9025-8

Braun, H. I. (2005). Using student progress to evaluate teachers: A primer on value-added models. *Educational Testing Service,* 1-20. Retrieved from http://files.eric.ed.gov/fulltext/ED529977.pdf

Chetty, R., Friedman, J., & Rockoff, J. (2014). Discussion of the American Statistical Association's Statement (2014) on using value-added models for educational assessment. *Statistics and Public Policy*, *1*(1), 111-113. doi:10.1080/2330443X.2014.955227

Coggshall, J. G., Ott, A., & Lasagna, M. (2010). *Convergence and contradictions in teachers' perceptions of policy reform ideas (Retaining Teacher Talent, Report No. 3).* Naperville, IL: Learning Point Associates and New York: Public Agenda. Retrieved from https://files.eric.ed.gov/fulltext/ED508143.pdf

Conley, S., & Glasman, N. (2008). Fear, the school organization, and teacher evaluation. *Educational Policy, 22*(1), 63-85. doi:10.1177/0895904807311297

Conley, S., Muncey, D. E., & You, S. (2005). Standards-based evaluation and teacher career satisfaction: A structural equation modeling analysis. *Journal of Personnel Evaluation in Education*, *18*(1), 39-65. Retrieved from https://eric.ed.gov/?id=EJ744204

U.S. Court of Appeals. (2015). *Cook et al. v. Bennett: Case: 14-12506*. Retrieved from http://media.ca11.uscourts.gov/opinions/pub/files/201412506.pdf

Creswell, J. W. (2015). *Qualitative inquiry & research design: Choosing among five approaches* (3 Ed.). Thousand Oaks, CA: Sage.

Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, *68*(4), 35-39. Retrieved from https://eric.ed.gov/?id=EJ913793

Darling-Hammond, L. (2015). Can value-added add value to teacher evaluation? *Educational Researcher, 44(2)*, 132-137. doi:10.3102/0013189X15575346

Donaldson, M. L., & Johnson, S. M. (2010). The price of mis-assignment: The role of teaching assignments in Teach for America teachers' exit from low-income schools and the teaching profession. *Educational Evaluation and Policy Analysis*, *32*(2), 299-323. Retrieved from http://journals.sagepub.com/doi/abs/10.3102

Doran, H. C., & Izumi, L. T. (2004). *Putting education to the test: A value-added model for California*. San Francisco: Pacific Research Institute.

Feeney, E. J. (2007). Quality feedback: The essential ingredient for teacher success. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *80*(4), 191-198. doi:10.3200/TCHS.80.4.191-198

Florida Department of Education [FLDOE]. (2017). *Performance evaluation: Florida's value-added models (VAM) frequently asked questions.* Retrieved from http://www.fldoe.org/teaching/performance-evaluation/

Florida Senate. (2018). *CS/HB 7069: Education.* Retrieved from https://www.flsenate.gov/Session/Bill/2017/7069

Florida Statute. (2011). *Florida Statute Section 1012.34- Personnel evaluation procedures and criteria.* Retrieved from http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&URL=1000-1099/1012/Sections/1012.34.html

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79*(4), 79-107. doi:10.1207/s15327930pje7904_5

Gearing, R. E. (2004). Bracketing in research: A typology. *Qualitative Health Research*, *14*(10), 1429-1452. doi:10.1177/1049732304270394

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing Company.

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). Evaluating teachers: The important role of value-added. *Brown Center on Education Policy at Brookings, 201*, 1-13. Retrieved from https://eric.ed.gov/?id=ED512829

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.tqsource.org/publications/ EvaluatingTeachEffectiveness.pdf

Goldhaber, D., & Hansen, M. (2010). Using performance on the job to inform teacher tenure decisions. *The American Economic Review*, *100*(2), 250-255. doi:10.1257/aer.100.2.250

Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy*, *1*(1), 28-39. doi:10.1080/2330443X.2013.856169

Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, *10*(1), 117-156. doi:10.1162/EDFP_a_00153

Haertel, E. (2013). Reliability and validity of inferences about teachers based on student test scores. *ETS,* 1-32. Retrieved from http://files.eric.ed.gov/fulltext/ED560957.pdf

Hanushek, E. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources, 14*(3), 351-388. Retrieved from http://www.jstor.org/stable/145575?seq=1#page_scan_tab_contents

Harris, D., & Anderson, A. (2012). Bias of public sector worker performance monitoring: Theory and empirical evidence from middle school teachers. *Association of Public Policy Analysis & Management*. Retrieved from https://appam.confex.com/appam/2012/webprogram/Paper2262.html

Hechinger Report. (2011). *Is the merit pay debate settled?* Retrieved from http://hechingerreport.org/is-the-merit-pay-debate-settled/

Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record*, *116*(1), 1-28. Retrieved from https://eric.ed.gov/?id=EJ1020230

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, *54*(5), 482-492. Retrieved from http://www.jstor.org/stable/20204940?seq=1#page_scan_tab_contents

Houston Federation of Teachers. (2017). *Houston Federation of Teachers Local 2415 et al v. Houston Independent School District*. Retrieved from: https://docs.justia.com/cases/federal/district-courts/texas/txsdce/4:2014cv01189/1174509/91

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics, 26*(1), 101-136. Retrieved from: http://ideas.repec.org/a/ucp/jlabec/v26y2008p101-136.html

Jacob, S. A., & Ferguson, S. P. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *The Qualitative Report*, *17*(42), 1-10. Retrieved from https://nsuworks.nova.edu/tqr/vol17/iss42/3/

Johnson, S. M., Papay, J. P., Fiarman, S. E., Munger, M. S., & Qazilbash, E. K. (2010). Teacher to teacher: Realizing the potential of peer assistance and review. *Center for American Progress*, 1-36. Retrieved from http://files.eric.ed.gov/fulltext/ED565879.pdf

Kim, J., & Suen, H. K. (2003). Predicting children's academic achievement from early assessment scores: A validity generalization study. *Early Childhood Research Quarterly, 18,* 547-566. doi:10.1016/j.ecresq.2003.09.011

LaParo, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research, 70*(4), 443-484. doi:10.3102/00346543070004443

LaVenia, M., Cohen-Vogel, L., & Lang, L. B. (2014). The common core state standards initiative: An event history analysis of state adoption. *American Journal of Education*, *121*(2), 145-182. doi:10.1086/679389

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Thousand Oaks, CA: Sage.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370-96. doi:10.1037/h0054346

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education*, *4*(4), 572-606. doi:10.1162/edfp.2009.4.4.572

McGaghie, W. C. (1991). Professional competence evaluation. *Educational Researcher*, *20*(1), 3-9. doi:10.3102/0013189X020001003

McGuinn, P. (2012). The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher-evaluation systems. *Center for American Progress,* 1-62. Retrieved from https://files.eric.ed.gov/fulltext/ED539744.pdf

Meisels, S. (2006). Accountability in early childhood: No easy answers. *Erikson Institute: Occasional Paper, 6,* 1-28. Retrieved from https://www.erikson.edu/research/accountability-in-early-childhood-no-easy-answers/

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. Hoboken, NJ: John Wiley & Sons.

Milanowski, A. T. (2005) Split roles in performance evaluation: A field study involving new teachers. *Journal of Personnel Evaluation in Education, 18*(3), 153-169. doi:10.1007/s11092-006-9017-0

Milanowski, A., Kimball, S., & White, B. (2004). The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites. *University of Wisconsin Working Paper Series TC*, *4*(01), 1-32.

Millman, J. (1997). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.

Moon, T. R., Brighton, C. M., Jarvis, J. M., & Hall, C. J. (2007). State standardized testing programs: Their effects on teachers and students. *The National Research Center on the Gifted and Talented, 1-260*. Retrieved from http://files.eric.ed.gov/fulltext/ED505375.pdf

Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives, 18*(23), 1-20. Retrieved from http://www.redalyc.org/html/2750/275019712023/

Papay, J. P., & Johnson, S. M. (2012). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *Educational Policy*, *26*(5), 696-729. doi:10.1177/0895904811417584

Ramirez, A. (2001). How merit pay undermines education. *Educational Leadership, 58(5),* 16-20. Retrieved from https://eric.ed.gov/?id=EJ622995

Reese, M., Gordon, S. P., & Price, L. R. (2004). Teachers' perceptions of high-stakes testing. *Journal of School Leadership*, *14*(5), 464-496. Retrieved from https://eric.ed.gov/?id=EJ852550

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, *8*(3), 299-311. doi:10.1007%2FBF00973726

Savin-Baden, M., & Major, C. H. (2013). *Qualitative research: The essential guide to theory and practice*. New York, NY: Routledge.

Sass, T. R., Semykina, A., & Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, *38*, 9-23. doi:10.1016/j.econedurev.2013.10.003

Sekaran, U. (2003). *Research methods for business: A skill building approach* (4th ed.). Hoboken, NJ: John Wiley and Sons.

Stake, R. (2005). *The art of case study research*. Thousand Oaks, CA: Sage Publications.

State Impact. (2014). *Senate Bill 736 - The student success act outlines how Florida teachers get paid.* Retrieved from http://stateimpact.npr.org/florida/tag/senate-bill-736/

Stitzlein, S. M., Feinberg, W., Greene, J., & Miron, L. (2007). Illinois project for democratic accountability. *Educational Studies*, *42*(2), 139-155. doi:10.1080/00131940701513235

Stronge, J. H. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Stronge, J. H., & Tucker, P. D. (1999). The politics of teacher evaluation: A case study of new system design and implementation. *Journal of Personnel Evaluation in Education, 13*(4), 339-359. doi:10.1023/A:1008105332543

Stronge, J., & Tucker, P. (2003). *Handbook on teacher evaluation: Assessing and improving performance*. Larchmont, NY: Eye on Education.

Teijlingen, E. R., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update,* 35, 1-4. Retrieved from http://aura.abdn.ac.uk/handle/2164/157

Tufford, L., & Newman, P. (2010). *Designing qualitative research*. Thousand Oaks, CA: Sage.

U.S. Department of Education. (2009). Race to the top program. *Executive Summary,* 1-15. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Van Wijk, E., & Harrison, T. (2013). Managing ethical problems in qualitative research involving vulnerable populations using a pilot study. *International Journal of Qualitative Methods, 12*(1), 570-586. doi:10.1177/160940691301200130

Watson, R., Atkinson, I., & Rose, K. (2007). Pilot studies: To publish or not. *Journal of Clinical Nursing, 16,* 619-620. doi:10.1111/j.136-2702.2006.01830.x

Weems, D. M., & Rogers, C. B. (2010). Are US teachers making the grade? A proposed framework for teacher evaluation and professional growth. *Management in Education*, *24*(1), 19-24. doi:10.1177/0892020609354959

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project*, 2-48. Retrieved from https://files.eric.ed.gov/fulltext/ED515656.pdf

Yin, R. K. (2013). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Appendix A: Semi-Structured Interview Questions

Demographic Questions

Age _____

Gender _____

Teaching experience (yrs.) _____

Ethnicity _____

Highest Degree earned _____

Grade teaching now _____

Annual teaching salary _____

Research Question-based Questions

1. Tell me about your district's teacher evaluation system.
2. How has the VAM adoption impacted your overall teaching practice?
3. Can you explain your belief in the VAM as a fair and accurate process in evaluating your teaching?
4. Have you received a VAM score on your evaluation and are you able to decipher it to be an accurate portrayal of your teaching?
5. What is your knowledge of the formula used to determine your VAM score?
6. Explain the extent to which the use of the VAM as half of your evaluation is a fair process.
7. Explain the extent to which the VAM has been useful to you in driving your instructional practice.

8. What is your perception of how the VAM is being used as part of your evaluation affects you personally?
9. What is your opinion of the VAM combined with your teacher evaluations results being used for the purpose of merit pay? Is it a fair system? Does it motivate you to improve? Does it inform you on the areas that you need to work on improving?
10. What is your understanding and knowledge of the VAM combined with the teacher evaluations being used for the purpose of job retention?
11. In what ways has the VAM evaluation system helped you do your job better?
12. Is there anything else you wish to say about VAM?

Appendix B: Florida's VAM Formula

In its most general formulaic form, the VAM can be represented mathematically as:

$$y_{ti} = \mathbf{X}_i\boldsymbol{\beta} + \sum_{r=1}^{L} y_{t-r,i}\gamma_{t-r} + \sum_{q=1}^{Q} \mathbf{Z}_{qi}\boldsymbol{\theta}_q + e_i$$

- y_{ti} is the observed score at time *t* for student *i*.
- $\mathbf{X}_i$ is the model matrix for the student and school level demographic variables.
- $\boldsymbol{\beta}$ is a vector of coefficients capturing the effect of any demographics included in the model.
- $y_{t-r,i}$ is the observed lag score at time *t-r* ($r \in \{1,2,\ldots,L\}$).
- $\boldsymbol{\gamma}$ is the coefficient vector capturing the effects of lagged scores.
- $\mathbf{Z}_{qi}$ is a design matrix with one column for each unit in *q* ($q \in \{1,2,\ldots,Q\}$) and one row for each student record in the database.

Appendix C: Recruitment Letter

Dear Teachers:

My name is Theresa Pascual. I am a reading and gifted consult teacher at GHS. Currently, I am a graduate student at Bethel University. I am conducting a research study about teachers' experiences of the value-added model used in their teacher evaluations. I would like to invite you to participate in my study. The purpose of my study is to explore teachers' professional and personal perceptions of the value-added model (VAM) to determine teachers' performance, compensation, and personnel consequences. This will be accomplished through interview questions that aim to measure teacher understanding and perceptions of the VAM. With the changes using the VAM in teacher evaluations in recent years, this study will look into the lived experiences of teachers and how they understand the evaluation system used in our district, the VAM in general, your personal VAM scores, and to determine if teachers believe errors exist in their actual performance appraisals.

The method I will be using for this research is a qualitative study which will consist of an interview. The interview will take place during a time and location of the participant's choosing. Participants will be asked a series of questions that relate to the VAM and their evaluations. This interview should not last longer than one hour. I will be making an audio recording of each interview in order to accurately capture the information discussed. I will also be taking handwritten notes during the interview. One year after the conclusion of this study, I will delete all audio recordings and destroy all hand-written notes. It is your decision whether or not to participate in this study.

The potential risks involved in this study are minimal. It is possible that participants may feel uncomfortable discussing their thoughts and feelings about VAM, or their experiences with specific teacher evaluations. I assure that measures will be taken to minimize any risks and/or discomforts associated with this study. I will request of every participant that all information discussed be held in strict confidence. In the final write-up of the study, I will use pseudonyms to protect the identity and privacy of all participants. If at any time a participant is not comfortable with the discussion involved in the interview, he/she may refuse to answer a question, or withdraw from the study all together.

Thank you so much for considering taking part in this study. Educational research is very important as we strive to continually improve our educational system. If you decide to participate, please write to me via email with your contact information and answers to the following demographic information: your age, gender, years of teaching experience, ethnicity, highest degree earned, grade/subject teaching now, tenured or annual contract, and annual teaching salary. I will then call you so that together we can establish a time, date, and location for the interview. Should you decide to participate, I will compensate you for your time with a $25.00 restaurant.com gift certificate.

Thanks again. I look forward to hearing from you!

Respectfully,

Theresa Pascual
Doctoral Student of Bethel University

Appendix D: Consent Form for Educational Research

Theresa Pascual, Doctoral Student of Bethel University

You are invited to participate in a study of the value-added model being used for a teacher's annual performance appraisals in Florida. I hope to learn teachers' professional and personal perceptions of the VAM to determine teachers' performance, compensation, and personnel consequences. This study will look into the lived experiences of teachers in understanding their VAM scores, which could determine if errors exist in their actual performance appraisals or if errors exist in a teacher's understanding of the VAM. The findings in this study could contribute to the research over the reliability and validity of VAM being used in teacher evaluation systems. You may feel you are taking a risk by revealing thoughts and attitudes about the VAM and/or your teacher evaluation. I assure all participants that measures will be taken to minimize any risks and/or discomforts associated with this study. Those measures include: participants will be informed that every effort will be made to keep their information confidential; I will request every participant to hold in strict confidence all given responses to questions; participants will be informed that the purpose of the study is to analyze the VAM in their teacher evaluation and will not affect their academic standing or placement; pseudonyms will be used for the school and all participants; codes will be used for participant responses so as to not reveal any private or personal information that can be linked to a particular person; all notes and recordings will be destroyed one year after the conclusion of this study; participants will be informed as to how their information will be used and disposed of; and prior to conducting the interview participants may refuse to respond to any questions or statements during the interview, and that they may withdraw from participation in the study at any point. Once the interview is transcribed, it will be sent back to participant through personal email, not school district email, for approval of accuracy. Finally, participants will be offered a $25.00 restaurant.com gift certificate, as inducements/rewards for participation in this study.

Any information I obtain in connection with this study that can be identified with you will remain confidential. In any written reports or publications, no one will be identified or identifiable and only aggregate data will be presented. I will be making an audio recording of each participant. The recordings will allow me to remain fully engaged in conversation, while taking journal notes. I will subsequently use the recordings to write transcripts, which will be coded and studied to determine trends in the findings. One year after the conclusion of this study, all audio recordings will be erased/deleted.

Your decision whether or not to participate will not affect your future relations with your school, Osceola County School District, or Bethel University in any way. If you decide to participate, you are free to discontinue participation at any time without affecting such relationships.

This research project has been reviewed and approved in accordance with Bethel's Levels of Review for Research with Humans.
You will be offered a copy of this form to keep.

Your signature indicates that you have read the information provided above and have decided to participate. You may withdraw at any time without prejudice after signing this form should you choose to discontinue participation in this study.

_____
Printed Name of Participant

_____ _____
Signature of Participant Date

_____ _____
Signature of Investigator Date

Appendix E: Informed Consent Process Outline

1. The principal investigator will contact administrators of participating schools through email requesting that the email invite be forwarded to their teachers asking for their permission to be interviewed for research.

2. Teachers that are interested in the study will email the principal investigator a letter of interest, along with their contact information and answers to all demographic questions.

3. The principal investigator will contact participants via telephone or email to explain the Informed Consent Form and instruct participants to return that form to the investigator via district office mail.

4. The principal investigator will collect all Informed Consent Forms through district office mail. Any unsigned Informed Consent Forms must be turned in prior to interviews.

Appendix F: Instructions Given to Participants Prior to Interview

I. Welcome, introduction of who I am, position, where I work
II. Explanation of the topic of study, purpose of the research
III. Review interview ground rules
    A. THERE ARE NO RIGHT OR WRONG ANSWERS
        i. Every person's experiences and opinions are important.
    B. WHAT IS SAID IN THIS ROOM STAYS HERE
        i. I want everyone to feel comfortable sharing when sensitive issues come up.
        ii. Confidentiality/Anonymity information will be reiterated.
    C. I WILL BE MAKING AN AUDIO RECORDING OF THE INTERVIEW
        i. I want to capture everything you have to say.
        ii. I don't identify anyone by name in the report. You will remain anonymous.
        iii. Hand-written journal notes will be taken during the recording.
    D. LENGTH OF THE INTERVIEW
        i. Interview should not take more than one hour to complete.
    E. INCENTIVE
        i. A $25.00 restaurant.com gift certificate will be issued at the end of the interview.
    F. EXCHANGE OF PERSONAL EMAILS
        i. At the end of interview, we will exchange personal emails for the purpose of sending the transcribed interview for approval of accuracy.

AT ANY TIME, YOU MAY CHOOSE TO END YOUR PARTICIPATION IN THIS INTERVIEW.

Title XLVIII **Chapter 1012** **View Entire Chapter**
K-20 EDUCATION CODE PERSONNEL

1012.34 Personnel evaluation procedures and criteria.—
(1) EVALUATION SYSTEM APPROVAL AND REPORTING.—
(a) For the purpose of increasing student academic performance by improving the quality of instructional, administrative, and supervisory services in the public schools of the state, the district school superintendent shall establish procedures for evaluating the performance of duties and responsibilities of all instructional, administrative, and supervisory personnel employed by the school district. The district school superintendent shall provide instructional personnel the opportunity to review their class rosters for accuracy and to correct any mistakes. The district school superintendent shall report accurate class rosters for the purpose of calculating district and statewide student performance and annually report the evaluation results of instructional personnel and school administrators to the Department of Education in addition to the information required under subsection (5).
(b) The department must approve each school district's instructional personnel and school administrator evaluation systems. The department shall monitor each district's implementation of its instructional personnel and school administrator evaluation systems for compliance with the requirements of this section.
(c) Annually, by February 1, the Commissioner of Education shall publish on the department's website the status of each school district's instructional personnel and school administrator evaluation systems. This information must include performance evaluation results for the prior school year for instructional personnel and school administrators using the four levels of performance specified in paragraph (2)(e). The performance evaluation results for instructional personnel shall be disaggregated by classroom teachers, as defined in s. 1012.01(2)(a), excluding substitute teachers, and all other instructional personnel, as defined in s. 1012.01(2)(b)-(d).
(2) EVALUATION SYSTEM REQUIREMENTS.—The evaluation systems for instructional personnel and school administrators must:
(a) Be designed to support effective instruction and student learning growth, and performance evaluation results must be used when developing district and school level improvement plans.
(b) Provide appropriate instruments, procedures, timely feedback, and criteria for continuous quality improvement of the professional skills of instructional personnel and school administrators, and performance evaluation results must be used when identifying professional development.
(c) Include a mechanism to examine performance data from multiple sources, including opportunities for parents to provide input into employee performance evaluations when appropriate.
(d) Identify those teaching fields for which special evaluation procedures and criteria are necessary.
(e) Differentiate among four levels of performance as follows:
1. Highly effective.
2. Effective.
3. Needs improvement or, for instructional personnel in the first 3 years of employment who need improvement, developing.
4. Unsatisfactory.
(f) Provide for training and monitoring programs based upon guidelines provided by the department to ensure that all individuals with evaluation responsibilities understand the

proper use of the evaluation criteria and procedures.

In addition, each district school board may establish a peer assistance process. This process may be a part of the regular evaluation system or used to assist employees placed on performance probation, newly hired classroom teachers, or employees who request assistance.

(3) EVALUATION PROCEDURES AND CRITERIA.—Instructional personnel and school administrator performance evaluations must be based upon the performance of students assigned to their classrooms or schools, as provided in this section. Pursuant to this section, a school district's performance evaluation system is not limited to basing unsatisfactory performance of instructional personnel and school administrators solely upon student performance, but may include other criteria to evaluate instructional personnel and school administrators' performance, or any combination of student performance and other criteria. Evaluation procedures and criteria must comply with, but are not limited to, the following:

(a) A performance evaluation must be conducted for each employee at least once a year, except that a classroom teacher, as defined in s. 1012.01(2)(a), excluding substitute teachers, who is newly hired by the district school board must be observed and evaluated at least twice in the first year of teaching in the school district. The performance evaluation must be based upon sound educational principles and contemporary research in effective educational practices. The evaluation criteria must include:

1. Performance of students.—At least one-third of a performance evaluation must be based upon data and indicators of student performance, as determined by each school district. This portion of the evaluation must include growth or achievement data of the teacher's students or, for a school administrator, the students attending the school over the course of at least 3 years. If less than 3 years of data are available, the years for which data are available must be used. The proportion of growth or achievement data may be determined by instructional assignment.

2. Instructional practice.—For instructional personnel, at least one-third of the performance evaluation must be based upon instructional practice. Evaluation criteria used when annually observing classroom teachers, as defined in s. 1012.01(2)(a), excluding substitute teachers, must include indicators based upon each of the Florida Educator Accomplished Practices adopted by the State Board of Education. For instructional personnel who are not classroom teachers, evaluation criteria must be based upon indicators of the Florida Educator Accomplished Practices and may include specific job expectations related to student support.

3. Instructional leadership.—For school administrators, at least one-third of the performance evaluation must be based on instructional leadership. Evaluation criteria for instructional leadership must include indicators based upon each of the leadership standards adopted by the State Board of Education under s. 1012.986, including performance measures related to the effectiveness of classroom teachers in the school, the administrator's appropriate use of evaluation criteria and procedures, recruitment and retention of effective and highly effective classroom teachers, improvement in the percentage of instructional personnel evaluated at the highly effective or effective level, and other leadership practices that result in student learning growth. The system may include a means to give parents and instructional personnel an opportunity to provide input into the administrator's performance evaluation.

4. Other indicators of performance.—For instructional personnel and school administrators, the remainder of a performance evaluation may include, but is not limited to, professional and job responsibilities as recommended by the State Board of Education or identified by the district school board and, for instructional personnel, peer

reviews, objectively reliable survey information from students and parents based on teaching practices that are consistently associated with higher student achievement, and other valid and reliable measures of instructional practice.
(b) All personnel must be fully informed of the criteria, data sources, methodologies, and procedures associated with the evaluation process before the evaluation takes place.
(c) The individual responsible for supervising the employee must evaluate the employee's performance. The evaluation system may provide for the evaluator to consider input from other personnel trained under subsection (2). The evaluator must submit a written report of the evaluation to the district school superintendent for the purpose of reviewing the employee's contract. The evaluator must submit the written report to the employee no later than 10 days after the evaluation takes place. The evaluator must discuss the written evaluation report with the employee. The employee shall have the right to initiate a written response to the evaluation, and the response shall become a permanent attachment to his or her personnel file.
(d) The evaluator may amend an evaluation based upon assessment data from the current school year if the data becomes available within 90 days after the close of the school year. The evaluator must then comply with the procedures set forth in paragraph (c).
(4) NOTIFICATION OF UNSATISFACTORY PERFORMANCE.—If an employee who holds a professional service contract as provided in s. 1012.33 is not performing his or her duties in a satisfactory manner, the evaluator shall notify the employee in writing of such determination. The notice must describe such unsatisfactory performance and include notice of the following procedural requirements:
(a) Upon delivery of a notice of unsatisfactory performance, the evaluator must confer with the employee who holds a professional service contract, make recommendations with respect to specific areas of unsatisfactory performance, and provide assistance in helping to correct deficiencies within a prescribed period of time.
(b)1. The employee who holds a professional service contract shall be placed on performance probation and governed by the provisions of this section for 90 calendar days following the receipt of the notice of unsatisfactory performance to demonstrate corrective action. School holidays and school vacation periods are not counted when calculating the 90-calendar-day period. During the 90 calendar days, the employee who holds a professional service contract must be evaluated periodically and apprised of progress achieved and must be provided assistance and inservice training opportunities to help correct the noted performance deficiencies. At any time during the 90 calendar days, the employee who holds a professional service contract may request a transfer to another appropriate position with a different supervising administrator; however, if a transfer is granted pursuant to ss. 1012.27(1) and 1012.28(6), it does not extend the period for correcting performance deficiencies.
2. Within 14 days after the close of the 90 calendar days, the evaluator must evaluate whether the performance deficiencies have been corrected and forward a recommendation to the district school superintendent. Within 14 days after receiving the evaluator's recommendation, the district school superintendent must notify the employee who holds a professional service contract in writing whether the performance deficiencies have been satisfactorily corrected and whether the district school superintendent will recommend that the district school board continue or terminate his or her employment contract. If the employee wishes to contest the district school superintendent's recommendation, the employee must, within 15 days after receipt of the district school superintendent's recommendation, submit a written request for a hearing. The hearing shall be conducted at the district school board's election in accordance with one of the following procedures:
a. A direct hearing conducted by the district school board within 60 days after receipt

of the written appeal. The hearing shall be conducted in accordance with the provisions of ss. 120.569 and 120.57. A majority vote of the membership of the district school board shall be required to sustain the district school superintendent's recommendation. The determination of the district school board shall be final as to the sufficiency or insufficiency of the grounds for termination of employment; or

b. A hearing conducted by an administrative law judge assigned by the Division of Administrative Hearings of the Department of Management Services. The hearing shall be conducted within 60 days after receipt of the written appeal in accordance with chapter 120. The recommendation of the administrative law judge shall be made to the district school board. A majority vote of the membership of the district school board shall be required to sustain or change the administrative law judge's recommendation. The determination of the district school board shall be final as to the sufficiency or insufficiency of the grounds for termination of employment.

(5) ADDITIONAL NOTIFICATIONS.—The district school superintendent shall annually notify the department of any instructional personnel or school administrators who receive two consecutive unsatisfactory evaluations. The district school superintendent shall also notify the department of any instructional personnel or school administrators who are given written notice by the district of intent to terminate or not renew their employment. The department shall conduct an investigation to determine whether action shall be taken against the certificateholder pursuant to s. 1012.795.

(6) ANNUAL REVIEW OF AND REVISIONS TO THE SCHOOL DISTRICT EVALUATION SYSTEMS.—The district school board shall establish a procedure for annually reviewing instructional personnel and school administrator evaluation systems to determine compliance with this section. All substantial revisions to an approved system must be reviewed and approved by the district school board before being used to evaluate instructional personnel or school administrators. Upon request by a school district, the department shall provide assistance in developing, improving, or reviewing an evaluation system.

(7) MEASUREMENT OF STUDENT PERFORMANCE.—

[1](a) The Commissioner of Education shall approve a formula to measure individual student learning growth on the statewide, standardized assessments in English Language Arts and mathematics administered under s. 1008.22. A third party, independent of the assessment developer, must analyze student learning growth data calculated using the formula and provide access to a data visualization tool that enables teachers to understand and evaluate the data and school administrators to improve instruction, evaluate programs, allocate resources, plan professional development, and communicate with stakeholders. The formula must take into consideration each student's prior academic performance. The formula must not set different expectations for student learning growth based upon a student's gender, race, ethnicity, or socioeconomic status. In the development of the formula, the commissioner shall consider other factors such as a student's attendance record, disability status, or status as an English language learner. The commissioner may select additional formulas to measure student performance as appropriate for the remainder of the statewide, standardized assessments included under s. 1008.22 and continue to select formulas as new assessments are implemented in the state system.

(b) Each school district may, but is not required to measure student learning growth using the formulas approved by the commissioner under paragraph (a).

[1](8) RULEMAKING.—The State Board of Education shall adopt rules pursuant to ss. 120.536(1) and 120.54 which establish uniform procedures and format for the submission, review, and approval of district evaluation systems and reporting requirements for the annual evaluation of instructional personnel and school administrators.

History.—s. 708, ch. 2002-387; s. 7, ch. 2004-255; s. 11, ch. 2004-295; s. 60, ch.

2006-74; s. 29, ch. 2008-108; s. 2, ch. 2011-1; s. 13, ch. 2014-23; s. 94, ch. 2014-39; s. 12, ch. 2015-6; s. 36, ch. 2017-116.
[1]Note.—Section 17, ch. 2011-1, provides that "[c]hapter 2010-279, Laws of Florida, does not apply to any rulemaking required to administer this act."

Appendix H: Sample VAM Report

Employee:

Location:

Evaluator Name:

Final Score: 3.4

Final Score Scale

Range 0.0 - 4.0				
Label	Highly Effective (HE)	Effective (E)	Developing (DV)	Unsatisfactory (U)
Details	3.5 - 4.0	2.0 - 3.49	1.5 - 1.99	0.0 - 1.49

Instructional Practice Score: 3.1

Instructional Practice Score Scale

Range 0.0 - 4.0				
Label	Highly Effective (HE)	Effective (E)	Needs Improvement (NI)	Unsatisfactory (U)
Details	3.5 - 4.0	2.5 - 3.49	1.5 - 2.49	0.0 - 1.49

Domain Count & Frequency:
Domain 1 score = 3

Label	Highly Effective	Effective	Needs Improvement	Unsatisfactory
Count	0	5	0	0

Domain 2 score = 3

Label	Highly Effective	Effective	Needs Improvement	Unsatisfactory
Count	0	4	0	0

Domain 3 score = 3

Label	Highly Effective	Effective	Needs Improvement	Unsatisfactory
Count	0	4	0	0

Instructional Status Score (90% of Instructional Practice Score): 3

Deliberate Practice Score (10% of Instructional Practice Score): 4

Deliberate Practice Details:

6. Identifying Critical Information

Deliberate Practice Scale:

(Current Year - Baseline Year) / Baseline Year = Percentage of Change

- The following chart shall be used to determine the point value to be assigned for the teacher's deliberate practice

percentage of change:

Deliberate Practice Score Point Values	
• 24% or less	1
• 25% - 49%	2
• 50% - 74%	3
• 75% or higher	4

Student Growth Score: 4

Student Growth Scale:

Weight: 35.0% | Range 0.0 - 4.0				
Label	Highly Effective (HE)	Effective (E)	Developing (DV)	Unsatisfactory (U)
Details	3.5 - 4.0	2.0 - 3.49	1.5 - 1.99	0.0 - 1.49

Professional & Ethical Behaviors Score: 3

Professional & Ethical Behaviors Score Scale

Weight: 20% | Range 0.0 - 4.0				
Label	Highly Effective (HE)	Effective (E)	Needs Improvement (NI)	Unsatisfactory (U)
Details	3.5 - 4.0	2.5 - 3.49	1.5 - 2.49	0.0 - 1.49

Evaluator's Comments:

Employee's Comments:

Note:

This document is not a representation of a final evaluation until both the Student Growth Value and Instructional Practice rating are applied to the final score.
By clicking the acknowledge below, I neither accept or deny.

ACCEPTED by !

Evaluatee

Evaluator