A STUDY OF ASSESSMENT RESULTS, TEACHER EFFECTIVENESS RATINGS, AND TEACHER PERCEPTIONS UNDER A VALUE-ADDED TEACHER EVALUATION MODEL

A DISSERTATION
SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF EDUCATION
BY
CHAD EDWARD MICHALEK
DR. SERENA SALLOUM – ADVISOR

BALL STATE UNIVERSITY
MUNCIE, INDIANA
DECEMBER 2014
A STUDY OF ASSESSMENT RESULTS, TEACHER EFFECTIVENESS RATINGS, AND TEACHER PERCEPTIONS UNDER A VALUE-ADDED TEACHER EVALUATION MODEL

A DISSERTATION

Submitted to the Graduate School

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Education

By

CHAD EDWARD MICHALEK

Dissertation Advisor: Dr. Serena Salloum

APPROVED BY:

____________________________________  __________________________
Dr. Serena Salloum, Committee Chairperson  Date

____________________________________  __________________________
Dr. John Ellis, Committee Member  Date

____________________________________  __________________________
Dr. Lynn Lehman, Committee Member  Date

____________________________________  __________________________
Dr. Ashley Donnelly, Committee Member  Date

____________________________________  __________________________
Dr. Robert Morris, Dean of Graduate School  Date

BALL STATE UNIVERSITY
MUNCIE, INDIANA
DECEMBER 2014
ABSTRACT

This dissertation addressed an evolving area of educational policy, value-added teacher evaluations. It investigated how teachers in one large urban Indiana school district perceived the use of value-added measures as part of an evaluation framework. A quantitative research design was utilized through various analyses of student assessment data in conjunction with teacher evaluation data. Additionally, teacher surveys were used to analyze teacher perceptions of the evaluation system.

Results suggest student mean test scores did not consistently improve under value-added teacher evaluations; however, in contrast student mean test score growth consistently improved at a statistically significant level under value-added teacher evaluations. Also, student achievement growth was most impacted by student previous year achievement score with a negative association, meaning that as a student’s previous year achievement score decreased achievement growth increased. Additionally, there was very poor consistency between teacher effectiveness ratings and principal-based observations. Finally, teachers generally perceived the new evaluation system negatively. These results were consistent with other research and have implications for teachers, administrators, parents, and policy makers.
ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my dissertation chair, Dr. Serena Salloum, for her guidance, patience, and support in completing my dissertation. I would also like to thank my dissertation committee members Dr. John Ellis, Dr. Lynn Lehman, and Dr. Ashley Donnelly for reviewing and shaping my research. I would like to thank my mother and father, Ed and Judy Michalek for always being loving and supportive parents and instilling in me the importance of education. I would like to thank my wife and best friend, Nikki Michalek, who has been my greatest champion in life. Finally, I would like to thank my sons, CJ and Alex Michalek. You have been and will continue to be my inspiration for everything I do in life. Every day you make me believe a better life is possible for all people. To all acknowledged here and those who have not been mentioned by name, I sincerely thank you for helping me complete my doctoral journey. I could not have done it without each and every one of you.
# TABLE OF CONTENTS

ABSTRACT ......................................................................................................................................................... iii

ACKNOWLEDGMENTS ........................................................................................................................................ iv

LIST OF TABLES ................................................................................................................................................ xii

INTRODUCTION .................................................................................................................................................. 14

  Problem Statement ......................................................................................................................................... 14
  School District ................................................................................................................................................. 17
  Evaluation Model ............................................................................................................................................. 17

Purpose Statement ............................................................................................................................................ 18

Research Questions ........................................................................................................................................... 19

Definition of Important Terms ......................................................................................................................... 20

Significance of the Study ................................................................................................................................. 21

Delimitations ..................................................................................................................................................... 22

  Time .............................................................................................................................................................. 22
  Location ....................................................................................................................................................... 23
  Teachers ......................................................................................................................................................... 23
  Schools ......................................................................................................................................................... 23

Assumptions ..................................................................................................................................................... 23

  Sample ......................................................................................................................................................... 23
  Fidelity ......................................................................................................................................................... 24
  Integrity ....................................................................................................................................................... 24
Summary .................................................................................................................................24

REVIEW OF THE LITERATURE ...........................................................................................25

Conceptual Framework ........................................................................................................26

Benefits of Value-Added Measures ......................................................................................27

Desired by the Public ........................................................................................................28

Increased Student Achievement .........................................................................................28

Predictable Student Achievement ......................................................................................29

Leveled Playing Field .........................................................................................................29

Increased Teacher Pay .......................................................................................................30

Increased Cost Efficiency ..................................................................................................30

Good Even when Bad ........................................................................................................31

Model Creator Claims .......................................................................................................32

Negatives of Value-Added Measures ..................................................................................33

Invalid .................................................................................................................................33

Unreliable ............................................................................................................................35

Data Issues ........................................................................................................................35

Floor and Ceiling Effects ..................................................................................................36

Lack of Transparency ........................................................................................................36

Misidentifying Effective Teachers .....................................................................................37

Claimed Usage in other Disciplines ..................................................................................38

Perverse Incentives ..........................................................................................................39

Not Utilized to Improve Instruction ..................................................................................39

Teaching to the Test ..........................................................................................................40
Decreased Motivation ................................................................. 41
Teacher Demoralization ............................................................ 41
Lack of Necessity ........................................................................ 42
Indiana’s Position on Value-Added Measures ............................... 42
Model Creator Claims ................................................................. 43
Desired by the Public ................................................................. 43
Invalid ....................................................................................... 44
Increased Cost Efficiency ............................................................ 44
Data Issues ................................................................................ 45
Mandated Failure ....................................................................... 45
Lack of Transparency ................................................................. 46
Less Teacher Collaboration .......................................................... 46
Unwillingness to Accept Student Teachers ................................. 47
Stunted Student Growth .............................................................. 47
Legal Issues ............................................................................... 49
Teacher Demoralization .............................................................. 52
Summary .................................................................................... 52
Relevance of Teacher Perceptions ............................................... 53
Relevance of Increased Student Achievement ............................ 55
Research Questions .................................................................... 55
Conclusion ................................................................................ 56
RESEARCH METHODS ................................................................ 58
VETCS ...................................................................................... 59
Sample .......................................................................................................................... 59
Variables .......................................................................................................................... 61
Measures ............................................................................................................................ 61
Statistical Relationships ...................................................................................................... 61
Limitations .......................................................................................................................... 62
VGCS .................................................................................................................................. 62
Sample ............................................................................................................................... 63
Variables ............................................................................................................................ 64
Measures ............................................................................................................................ 64
Statistical Relationships ...................................................................................................... 65
Limitations .......................................................................................................................... 65
VMRS .................................................................................................................................. 66
Sample ............................................................................................................................... 66
Variables ............................................................................................................................ 67
Measures ............................................................................................................................ 67
Multiple Regression ............................................................................................................ 70
Statistical Relationships ...................................................................................................... 71
Limitations .......................................................................................................................... 71
VCCS .................................................................................................................................. 71
Sample ............................................................................................................................... 71
Variables ............................................................................................................................ 72
Measures ............................................................................................................................ 72
Statistical Relationships ...................................................................................................... 73
Replication Studies .................................................................................................................................. 111
Departed Educator Interviews ............................................................................................................ 111
Comparison to Enhanced VAM .......................................................................................................... 111

REFERENCES ........................................................................................................................................ 112

APPENDIX A: SURVEY RECRUITMENT LETTER .............................................................................. 124
APPENDIX B: TEACHER PERCEPTIONS OF THE USE OF VALUE-ADDED MEASURES IN TEACHER EVALUATIONS SURVEY ......................................................................................... 126
APPENDIX C: ISTEP+ CUT SCORES .................................................................................................... 135
APPENDIX D: ADDITIONAL TABLES OF SURVEY PARTICIPANT DEMOGRAPHICS 136
# LIST OF TABLES

Table 3.1: VETCS Sample Grouping..................................................................................60
Table 3.2: VETCS Test Statistic and Statistical Procedure to be Used..............................61
Table 3.3: VGCS Sample Grouping...................................................................................64
Table 3.4: VGCS Test Statistic and Statistical Procedure to be Used.................................65
Table 3.5: VMRS Test Statistic and Statistical Procedure to be Used.................................69
Table 3.6: VCCS Test Statistic and Statistical Procedure to be Used.................................73
Table 4.1: t-Test for Equality of Means for ELA, Math, Science, and Social Studies Scale
Scores..................................................................................................................................80
Table 4.2: t-Test for Equality of Means for Growth in ELA and Math Scale
Scores..................................................................................................................................82
Table 4.3: Multiple Regression Model Summary for Growth in ELA and Math Scale
Scores..................................................................................................................................83
Table 4.4: Multiple Regression Model Summary for 4th Grade Growth in ELA and Math
Scores..................................................................................................................................84
Table 4.5: Multiple Regression Model Summary for 5th Grade Growth in ELA and Math
Scores..................................................................................................................................85
Table 4.6: Multiple Regression Model Summary for 6th Grade Growth in ELA and Math
Scores..................................................................................................................................86
Table 4.7: Multiple Regression Model Summary for 7th Grade Growth in ELA and Math Scores…………………………………………………………………………………………87
Table 4.8: Multiple Regression Model Summary for 8th Grade Growth in ELA and Math Scores…………………………………………………………………………………………88
Table 4.9: Survey Question: For each statement below, please indicate your level of agreement…………………………………………………………………………………………92
Table 4.10: Survey Question: How familiar are you with Public Law 90 that required the implementation of a new teacher evaluation system starting in the 2012-13 school year?…………………………………………………………………………………………93
Table 4.11: Survey Question: How have you become familiar with the requirements of Public Law 90? (Select all that apply)…………………………………………………………………………………………94
Table 4.12: Survey Question: Which stakeholder groups that you know of have been or will be a part of your district's VAM-based teacher evaluation development process? (Select all that apply)…………………………………………………………………………………………94
Table 4.13: Survey Question: For each statement below, please indicate your level of agreement…………………………………………………………………………………………95
Table 4.14: Survey Question Responses Regarding Teaching Profession Perspectives………………………………………………………………………………………………………98
Table 4.15: Survey Question Responses Regarding Perspectives on Cheating……………………………………………………………………………………………………99
CHAPTER 1

INTRODUCTION

Problem Statement

In the fall of 2012, Indiana schools began transitioning to a new teacher evaluation model that included value-added measures (VAM). These measures are a collection of statistical formulas and techniques that try to derive teacher effects on student performance. Essentially, VAMs are believed to, “help to evaluate the knowledge that teachers add to student learning as students progress through school” (Amrein-Beardsley, 2008, p. 65). The approach Indiana has mandated uses:

- student assessment results from statewide assessments for certificated employees whose responsibilities include instruction in subjects measured in statewide assessments,
- methods for assessing student growth for certificated employees who do not teach in areas measured by statewide assessments, and student assessment results from locally developed assessments and other test measures for certificated employees whose responsibilities may or may not include instruction in subjects and areas measured by statewide assessments. (Indiana Code 20-28-11.5)

As noted above, Indiana’s new evaluation model at the time of this writing required the use of growth measures to evaluate teacher effectiveness which is a form of VAM. Specifically, Indiana used a growth model dimension that:
uses standardized testing scores to create “academic peers” to compare student progress. Students are academic peers if (a) they are in the same grade level, (b) they are taking the same test, and (c) they have the exact same score. As a two-year process, students are placed into academic peer groups in year one. In year two, the academic peer group’s scores are compared on a normal distribution, represented in percentiles. (Whiteman, Shi, & Plucker, 2011, p. 12)

Any teacher that negatively affects student achievement or growth as indicated by VAM, through the interpretation and guidance of the Indiana Department of Education, cannot receive a pay increase and could be terminated depending on teaching experience and past evaluation ratings (Indiana Department of Education, 2012). VAM is related to test based-accountability which is a result of the No Child Left Behind Act of 2001 (NCLB). NCLB mandated standardized testing to report on student progress through the use of school-based value-added models for accountability purposes (Olson, 2004). The switch to teacher-based value-added models was a result of the 2012 federal Race to the Top grant program (RTT) which required the use of teacher affected student growth measures to document the number of effective and highly effective teachers in a school (U.S. Department of Education, 2012). Additionally, it is believed that VAM can, “capture how much students learn during the school year, thereby putting teachers on a more level playing field as they aim for tenure or additional pay” (David, 2010, p. 81). However, there is already some concern from educators and researchers that these claims may not be true, especially considering Indiana’s interpretation of VAM (Cavanaugh, 2011; Cole, Robinson, Ansaldo, Whiteman, & Spradlin, 2012; Whiteman, Shi, & Plucker, 2011).

With the implementation of any new evaluation system, the evaluated individuals will likely be concerned. In fact the use of terms such as performance pay and merit pay have been
shown to provoke highly charged responses in teachers (Eckert & Dabrowski, 2010). However, there is no systematic, empirical evidence whether or not Indiana teachers are concerned with the new evaluation system and whether or not the value-added portion of the system is the real cause for apprehension. While the idea is new to Indiana, similar models have been used in other states such as Tennessee, Pennsylvania, North Carolina, as well as other local districts across the United States (Amrein-Beardsley, 2008). Research has been conducted on many facets of value-added measures in education that are potential causes of concern and might explain potential teacher distress. Tying teacher pay and student performance on high stakes tests is comparable to distributing an experimental drug before it has been adequately tested (Farmer, 2009). Value-added models in other states have been criticized for a variety of reasons including a lack of external review, a lack of transparency, issues with missing data, and a lack of consideration for student background data (Amrein-Beardsley, 2008). Additionally, tests used in value-added models have been criticized for not having sufficient stretch which means that the tests do not have ceiling effects (Koedel & Betts, 2008). If a test has a ceiling effect then true student educational attainment is constrained, thus limiting the impact a teacher can have on student growth. All of these problems are cause for concern. However, it is unknown whether teachers are aware of these issues, or if there are other issues that are more troubling to these evaluated teachers.

This research addressed an evolving area of educational policy: teachers’ perceptions of value-added teacher evaluations. It investigated how teachers in Indiana perceived the use of value-added measures as part of an evaluation framework. A quantitative research design was utilized with analysis of student test scores and Teacher Effectiveness Ratings (TER) under a value-added teacher evaluation model (VAMIN) used in the Value-Added Model School District
(VAMSD). VAMSD and VAMIN are generic names given to an actual school district and its value-added teacher evaluation to provide anonymity to the teachers and students. Additionally, an analysis of related teacher survey questions was conducted.

The results of this project have the potential to influence future educational policy, improve teacher-administrator relationships, and improve student educational outcomes by detailing the perceptions of teachers and statistical analysis of impacted student test scores.

**School District**

VAMSD is a single school district located in northern Indianapolis, Indiana. It has a dynamic community within an urban setting with broad diversity in cultures, religions, ethnic groups, races, and socioeconomic levels. The school district currently serves approximately 11,000 students. VAMSD has been an educational leader in Indiana for the past forty years by offering a comprehensive educational curriculum with special activities and programs geared to provide enrichment, exploration, and instructional support for students. Many former students and faculty have gone on to take prestigious positions in education, industry, and politics. VAMSD’s demographics made it an excellent choice for studying the impacts of VAM on teachers because the district has a rich mix of beliefs, cultures and economic diversity that allowed factors influencing VAM success or failure to be more easily identified.

**Evaluation Model**

VAMSD uses a teacher evaluation model that is comprised of traditional teacher evaluation components and VAM components. The traditional teacher evaluation components are formal and informal classroom observations by administrators in conjunction with a rubric that covers several qualitative best practice teaching domains. The VAM components take into account school and district A-F accountability measures and student growth percentiles (SGP).
The SGP components account for approximately 13.5 percent of a teacher’s total or summative evaluation. Summative evaluations are then converted into ratings of 1 through 4 representing ineffective and highly effective respectively.

Teachers in grades four through eight have their SGP calculated by the Indiana Department of Education (IDOE) in the form of TER also referred to as Individual Growth Models (IGM). The model used for TER calculates growth over a two year period. Student scores in year one are used to group all students with the same year one score. Each student’s year two score is used to calculate student growth in the form of a percentile of the students across the state who had the same starting score in math and English. All students, who attended school at least 162 days during the school year, growth percentiles in a teacher’s class are ordered from least to greatest and the median is calculated. Then interval estimates which are essentially confidence intervals of one standard error above and below the median are calculated to create upper and lower bounds. If a teacher’s upper bound is below 35 then he or she is rated a 1. If a teacher’s upper bound extends above 35 and at or below 50, he or she is rated a 2. If a teacher’s upper bound extends above 50 and at or below 65, he or she is rated a 3. If a teacher’s lower bound extends above 65, he or she is rated a 4.

**Purpose Statement**

The purpose of this dissertation was twofold. The first purpose was to understand the impact of VAMIN on student achievement as indicated through a pre-VAMIN and post-VAMIN comparison of student achievement and growth. The second purpose was to investigate, relationally, if the method had any externalities that specifically impact teachers.

This research provided insight into whether or not VAMIN played a role in improving student achievement and student growth on ISTEP+ assessments. This is very important because
it helps to answer one of the most fundamental questions about any educational model, does it improve student outcomes.

Also, the research gave voice to teachers on an important and pressing matter of education policy. Authentic teacher perceptions have the ability to contextualize the limitations and hopes of using these models for evaluation purposes. This is extremely important because, as Elmore (2004) has indicated, teachers routinely have policy forced on them without teacher input. They are then expected to implement the policy. Thus, effective education policy implementation is very dependent on teacher buy-in. Additionally, this research offered insights into teacher beliefs and emotions that could be used to shape policy. This research helped provide a better understanding of the fears, reluctances, hopes, and aspirations teachers feel about the utilization of this evaluation technique in an effort to find possible areas of improvement that allow the model to more effectively evaluate teacher performance.

**Research Questions**

The research questions addressed in this dissertation are:

- Does the use of value-added teacher evaluations impact student achievement scores in VAMSD?
- Does the use of value-added teacher evaluations impact student achievement growth in VAMSD?
- If student achievement or student achievement growth increases under value-added teacher evaluations, is the value-added teacher evaluation model the factor responsible for the increase?
- Is there consistency between traditional teacher evaluations and value-added teacher evaluations in VAMSD?
• In what ways has the use of value-added teacher evaluation models impacted educators in VAMSD?

**Definition of Important Terms**

*Ceiling effect* refers to an upper limit score on a test that bounds measures of student achievement.

*Concurrent validity* refers to a measure of correlation between multiple measures to demonstrate whether or not different measures yield similar results.

*Construct under-representation (CUR)* refers to a test not having enough items dealing with the intended criterion to be measured on the test to measure the criterion.

*Construct-irrelevant variance (CIV)* refers to a test that requires skills that it does not intend to measure.

*Education value-added assessment system (EVAAS)* refers to a value-added method that calculates a teacher’s effect on student performance by tracking student progress on standardized assessments over a student’s academic career.

*Floor effect* refers to a lower limit score on a test that bounds measures of student achievement.

*Generalized cumulative effects model (CEM)* refers to an economic framework that addresses social inputs as part of a value-added model.

*Human capital theory* refers to a theory in which teachers create economic inputs for the nation in the form of students educated as future employees.

*Individual Growth Models (IGM)* refers to the name of the model used to calculate teacher effectiveness ratings in Indiana.

*Stretch* refers to a test that does not have a ceiling effect.
Student growth percentiles (SGP) refers to the model used for teacher effectiveness calculations in Indiana. It calculates growth over a two year period. Student scores in year one are used to group all students with the same year one score. Each student’s year two score is used to calculate student growth in the form of a percentile in comparison to the students across the state who had the same starting score in math and English.

Teacher effectiveness ratings (TER) refer to value-added scores in the form of student growth percentiles calculated by the Indiana Department of Education for Indiana teachers in grades four through eight.

Value-added measures (VAM) refers to a collection of statistical formulas and techniques that try to derive teacher effects on student performance.

Value-added model school district (VAMSD) refers to the generic name for the school district studied in this dissertation to provide anonymity to teachers and students.

Value-added teacher evaluation model (VAMIN) refers to the generic name for the value-added teacher evaluation model studied in this dissertation to provide anonymity to teachers and students.

Significance of the Study

This study is important because it answered two pressing questions: is there evidence that the use of VAMIN increases student achievement, and as Indiana has invested vast resources in this new evaluation system, how do teachers perceive the use of VAMIN?

Improving student achievement is one of the most important functions of an educational system. Thus, efforts to improve student achievement are rigorously pursued by educators. If value-added teacher evaluations do not help to improve student achievement, then they should be discontinued or altered so that student achievement does improve.
Additionally, teacher perceptions could have wide ranging effects on the education system itself. If teachers feel too denigrated by the new system or believe it has no credibility, there could be a mass exodus from the profession that would have many unintended political and sociological consequences. Additionally, if teachers feel higher levels of stress than they did before the implementation, there could be increased health care costs and sick days taken by teachers. Also, the increased pressure of attaining high value-added measure scores could cause teachers to attend school when they are ill. This could result in additional student illness and increased absenteeism among students. These issues could potentially demonstrate resulting hidden inefficiencies in the educational and political structures that would warrant further investigation in the use of value-added measures in evaluation models. Additionally, the use of VAM could show increased student achievement or a narrowing of the achievement gap which might mitigate teacher apprehension. Ultimately, this study could begin to answer the question of whether or not the implementation of this program is worth the cost, economically and politically (Taylor & Bogotch, 1994).

Delimitations

The following areas represent the restrictions of this study.

Time

This study is based on ISTEP+ data from the 2010-2011, 2011-2012, and 2012-2013 school years to coincide with the first year of VAMIN implementation and the previous non-VAMIN year to model the effect on student achievement. Additionally, survey data is based on VAMSD teachers who taught during the 2013-2014 school year as this expanded the potential for survey participants. The 2013-2014 school year is the second required year of VAM-based teacher evaluations in VAMSD. Thus, these teachers have both taught and received teaching
evaluations under VAMIN. Additionally, it was the first year in which the Indiana Department of Education had calculated TERs for school districts.

**Location**

This study focused solely on Indiana public schools represented by VAMSD, which is not representative of the state of Indiana. However, the diversity represented in the school district’s demographics more closely resembles the demographics of the United States. Thus, while the results do not necessarily generalize to Indiana, they might have broader appeal nationally.

**Teachers**

Those surveyed in this study consisted of licensed Indiana public school teachers in grades K-12 who taught during the 2013-2014 school year in VAMSD. While there are teachers who have left VAMSD over the last school year, using current teachers as survey participants likely yielded the best results because current teachers have the best insight into why teachers left the district and whether or not VAM had any impact on the decision to leave the district.

**Schools**

VAMSD schools were analyzed in this study in terms of their pre and post-VAMIN ISTEP+ scores in grades 3-8 as well as teacher perceptions on using VAM-based teacher evaluations. Also, possible externalities were investigated for VAMSD.

**Assumptions**

This study operated under the following assumptions.

**Sample**

The sample of VAMSD teachers who used the VAMIN teacher evaluation method is representative of teachers in large urban school districts in Indiana.
**Fidelity**

VAMSD and VAMIN school administrators and teachers implemented and utilized the VAMIN teacher evaluation model with fidelity.

**Integrity**

Responses received from teachers accurately reflected their professional opinions. Data provided by the Indiana Department of Education and VAMSD was accurate and error-free.

**Summary**

This research is an educational administration dissertation that addresses an evolving area of educational policy. It relationally investigated whether or not the VAMIN method raised student achievement and investigated how teachers in Indiana perceive the use of VAM as part of an evaluation framework. The purpose of this research was to review what fears, reluctances, hopes, and aspirations, if any, teachers have about the utilization of this evaluation technique in an effort to identify possible areas of improvement that would allow the model to more effectively evaluate teacher performance. The results of this project have the potential to influence future educational policy, improve teacher-administrator relationships, and improve student educational outcomes.
CHAPTER 2

REVIEW OF THE LITERATURE

The purpose of this literature review is twofold. The first and main purpose of this review is to establish that there is an important gap in published research about the effects of value-added evaluations on school teachers, and that such a project is a worthwhile. The research is first analyzed by perspective, pro-value-added measures versus anti-value-added measures. Then, the research is analyzed by themes such as benefits of VAM and problems of VAM.

The second purpose of this literature review is to identify potential perspectives teachers evaluated through VAM may hold. Teacher perspectives are incredibly important as teachers are being held solely responsible under the model for creating human capital through student development. If teachers are negatively impacted by the new system, psychologically or emotionally, this could impact the creation of capital which would impede the model’s sole purpose. The most important themes are listed as positives and negatives and were believed to be ideas likely addressed by teachers during surveys.

Value-added is an important and popular economic idea that is relevant in many occupational fields outside of education such as finance, health care, and government. This literature review only focused on VAM in terms of an educational perspective. Additionally, VAM in education are currently used to evaluate administrators and schools as well as teachers in primary, secondary, and post-secondary institutions. Thus, research addressing these issues is
available. This literature review does not include research focused on evaluating administrators or schools unless it addresses teacher evaluation as well. It includes research involving primary, secondary, and post-secondary teacher evaluation, as all areas are relevant to teacher evaluation. These decisions were made so that the scope of this work is limited to value-added teacher evaluations only.

This chapter is important in that it was the preliminary step in determining whether studying teacher perceptions about value-added evaluations was a relevant research endeavor. This literature review demonstrates that there is a need for this project. Ultimately it supports the thesis that it is imperative to study teacher perceptions as Indiana has invested vast resources and politic capital in the implementation of this new evaluation system based on value-added measures.

**Conceptual Framework**

Determining how valuable a teacher is has been an educational issue for many years. In fact, the idea of teachers creating economic inputs for the nation in the form of students educated as future employees was posited by Becker (1964) in his human capital theory. This theory undergirds the current educational policy environment dealing with VAM (Spring, 2008; Whiteman, Shi, & Plucker, 2011). Human capital theory represents future economic opportunity as dependent upon the abilities of the workforce derived directly from educational success. While human capital theory is appropriate, it views all educational inputs as responsible for “adding value” to the workforce through knowledge. This framework is limited by neglecting social inputs as factors in producing effective employees. Essentially, the model holds all teachers solely responsible for the quality of potential employees without considering important factors such as health, wealth, and access.
A framework that addresses social inputs as part of value-added is the generalized cumulative effects model (CEM) which includes all relevant past student, family, and educational inputs (Boardman & Murnane, 1979; Guarino, Reckase, & Woolridge, 2012; Hanushek, 1979, 1986; Todd & Wolpin, 2003). CEM uses quantitative models to take into consideration a variety of variables such as socioeconomic status, parental level of education, as well as many others. However, Indiana’s VAMIN model that will be studied in this piece does not take into account any input variables other than the teacher. Thus, the conceptual framework for this project is teacher as producer, a variant of the generalized cumulative effects model that views the teacher as the single input of a production function (Raudenbush & Bryk, 2002; Schmitz & Raymond, 2008). In terms of human capital theory, the teacher is solely attributable to adding value to the workforce through student assessment mastery and growth.

**Benefits of Value-Added Measures**

Most of the research conducted on VAM has focused on validity (Amrein-Beardsley, 2008; Baker, et al., 2010; Broatch & Lohr, 2012; van de Grift, 2009; Jacob & Lefgren, 2007) and reliability (Kane & Staiger, 2008; van de Grift, 2009) as a measure of teacher performance. Additionally, researchers have investigated the effects of student motivation and assessment (Anderman, Anderman, Yough, & Gimbert, 2010). Misco concluded that “interpretation of value-added data does not account for school role in larger societal milieu, and it sublimes political forces and other contextual features, thereby attributing disproportional weight to schools” (2008, p. 13). All of these results explain important aspects of whether value-added modeling is effective in evaluating teachers, but none examine how VAM impact the psyche and emotional makeup of teachers. In fact, little if any research has been conducted on how teachers perceive the use of these measures as a part of an evaluation. This dissertation presents the
opportunity to shed light onto how teachers are affected by the implementation of these instruments.

While these suggestions have not been validated by research, some available studies indicate positive aspects of using VAM to evaluate teachers.

**Desired by the Public**

The first and maybe most important issue related to the use of VAM to evaluate teachers is that the public largely supports this type of accountability. In a 2007 Phi Delta Kappa/Gallup Poll, citizens were given the following question:

One way to measure a school’s performance is to base it on the percentage of students passing the test mandated by the state at the end of the school year. Another way is to measure the improvement students in the school made during the year. In your opinion, which is the best way to measure the school’s performance—the percentage passing the test or the improvement shown by the students? (Rose & Gallup, 2007, p. 35)

The response to this question was overwhelmingly in favor of measuring student growth with eighty-two percent of the respondents selecting the second choice (Rose & Gallup, 2007).

**Increased Student Achievement**

Another important result of VAM is that these models have the potential to increase student achievement. This in and of itself should validate the use of these measures, because, ultimately, the point of all facets of education should be to raise student achievement and narrow the achievement gap. Demie’s (2003) research showed that the utilization of VAM by teachers to self-evaluate resulted in higher standardized test score gains than in schools that did not self-evaluate. In another study by Aaronson, Barrow, and Sander (2007), it was shown that a strong value-added teacher can raise student achievement. In fact, a classroom teacher that is one
standard deviation better than the average teacher, in terms of his or her value-added quality, will raise student scores that he or she teaches by one-fifth of the average yearly gains of all students. Additionally, moving an average teacher to the 84th percentile of the value-added spectrum would equate to increasing student performance from the 50th percentile to the 58th percentile (Hanushek & Rivkin, 2008). While these studies present VAM as a way to increase student achievement, they also are supporting improved teacher practices. The study by Demie (2003) represents the positive impact of teacher self-reflection while the studies by Aaronson, Barrow, and Sander (2007) and Hanushek and Rivkin (2008) illustrate that quality teachers raise student academic achievement.

**Predictable Student Achievement**

Another important aspect of VAM is that they can be used to predict student achievement. In research by Konstantopoulos and Sun (2012), they demonstrated that not only can VAM be used to predict student achievement in the current year but additionally these models can predict achievement in subsequent years. This has positive implications for student scheduling because the models have the potential to reduce administrator guesswork in student placement. Traditionally, administrators place students by self-selection or teachers’ perception of student ability. If models could be used to reliably predict how students will perform in the future, students could be placed in an appropriate class to aid in student academic development.

**Leveled Playing Field**

Using student standardized test scores to evaluate teachers is problematic because students entering a teacher’s classroom at the beginning of the year have multiple starting points. At the end of the year when students take standardized exams, the results are a one-time snapshot that only shows the finish line not the starting point of what a student has learned over the course
of the year. Thus, growth models have been included in VAM in which students are given a pretest and a posttest or a prior year’s test is compared to the current year’s test. This allows student growth over the year to be demonstrated. This ultimately acts to equal the playing field for teachers by negating the benefits or detriment of student starting points (Ballou, 2002). Under this methodology, teachers no longer need to worry about a student’s past academic success or failure in terms of their personal teaching evaluation.

**Increased Teacher Pay**

Proponents of value-added teacher evaluations claim that teachers should be paid more for the value they add to student achievement. However, it is often difficult to determine how valuable a teacher is. In research by Chingos and West (2012) it was shown through the use of VAM that teachers who have higher value-added scores earn more than low value-added teachers when they leave the teaching profession. This ultimately demonstrates that high value-added teachers should be paid more because they can make more money outside of education if they choose to leave.

**Increased Cost Efficiency**

Another potential advantage of using VAM to evaluate teachers is cost efficiency. Traditional measures of teacher evaluation usually center around classroom observations by school administrators. These evaluations take time to conduct and additional time to write up. VAM do not necessitate the same time investment as traditional hiring and evaluation practices. Lefgren and Sims (2012) showed that using simple one or two year VAM could more efficiently evaluate teachers as compared to traditional methods of teacher evaluation by reducing teacher evaluation to a simple process of inputting student data into a computer model that calculates results to rate teacher effectiveness. This coupled with research conducted by Staiger and
Rockoff (2010) could save school districts hundreds of thousands of dollars. This kind of savings, if true, would be a huge incentive to tax-payers.

**Good Even when Bad**

Even though many benefits of value-added have been listed, there are critics that claim these measures are flawed as will be investigated in subsequent sections of this review. Sampling or the selection of students included in an educator’s VAM, is an area that has brought criticism to the use of these measures. However, Winters, Dixon, and Greene (2012) showed that even when sampling issues are present in value-added evaluation models, failing to account for it does not substantially bias the estimation. Additionally, in a recently released study by the Bill and Melinda Gates Foundation (2012) on the Measures of Effective Teaching (MET) project, the authors claim that VAM should be included as a measure of teacher effectiveness despite two caveats:

First, a prediction can be correct on average but still be subject to prediction error. For example, many of the classrooms taught by teachers in the bottom decile in the measures of effectiveness saw large gains in achievement. In fact, some bottom decile teachers saw average student gains larger than those for teachers with higher measures of effectiveness. But there were also teachers in the bottom decile that did worse than the measures predicted they would. Anyone using these measures for high stakes decisions should be cognizant of the possibility of error for individual teachers. (Bill and Melinda Gates Foundation, 2012 p. 3)

Second, as a practical matter, we could not randomly assign students or teachers to a different school site. As a result, our study does not allow us to investigate the validity of the measures of effectiveness for gauging differences across schools. The process of
student sorting across schools could be different than sorting between classrooms in the same school. Yet school systems attribute the same importance to differences in teachers’ measured effectiveness between schools as to within-school differences. Unfortunately, our evidence does not inform the between-school comparisons in any way. (Bill and Melinda Gates Foundation, 2012 p. 5)

This ultimately suggests that there are some flaws in the use of VAM, but those problems should not prohibit the models from yielding beneficial information. This seems very Machiavellian, in that, even if these measures might misrepresent individual teacher effectiveness that could lead to job termination, that their use is justifiable because the results still give a somewhat sound interpretation of teacher effectiveness.

**Model Creator Claims**

The creators of VAM have made many claims that VAM have great educational benefits. For example, the Education Value-Added Assessment System (EVAAS) is considered to be the most recognized, widely implemented, and model upon which most VAM systems are based (Amrein-Beardsley, 2008). EVAAS is a value-added method that calculates a teacher's effect on student performance by tracking student progress on standardized assessments over a student’s academic career. The model has had the following claims made about its utilization:

- “It’s unimpaired by students’ backgrounds (race and levels of poverty), which distort all other analyses of student test score data” (Amrein-Beardsley, 2008, p. 66).
- “It’s not compromised by issues of missing data” (Amrein-Beardsley, 2008, p. 66).
- “It’s suitable for wide implementation across states because the software for processing data permits large-scale analyses” (Amrein-Beardsley, 2008 p.66).
- “Educational findings that were invisible in the past are now readily apparent” (Sanders
& Horn, 1994, p. 310).

“Both accelerators and impediments to sustained academic growth can be measured in a fair, objective and unbiased manner” (SAS, 2007).

“Without this information, educational improvement efforts cannot address the real factors that have been proven to have the greatest effect on student learning” (Sanders & Horn, 1998, p. 256).

“NCLB raises the academic standard for all kids, while the value-added approach is going beyond that and attempting to reach an even higher standard for individuals” (Sanders, 2004).

“There may be some important unintended consequences of this legislation if states do not go beyond the No Child Left Behind AYP (Adequate Yearly Progress) requirement and adopt a value-added model” (Sanders, 2003, p. 1).

These claims made by the model’s creators are very promising and demonstrate an improved method of evaluating teachers. However, these statements have never been validated by their authors through external peer-reviewed research.

**Negatives of Value-Added Measures**

While research noted above showed positive aspects of VAM used to evaluate teachers, there is also research depicting the negatives of value-added teacher evaluations as indicated below.

**Invalid**

The most important issue in adopting a VAM-based model is whether or not the test used as the basis for the evaluation is valid. This means: does the test measure what it is supposed to measure? In the case of value-added teacher evaluations, the question of validity is whether
these models truly measure and identify which teachers are effective and ineffective at raising student achievement and if those measures are solely attributable to the teacher. This has become the main criticism against VAM as researchers have questioned the model’s ability to measure growth especially at the tails of student achievement distributions (Amrein-Beardsley, 2008). Thus, if these models cannot assess growth for low and high-ability students, why would anyone want to make decisions based on the results? Broatch and Lohr (2012) showed that prediction by value-added evaluation models is problematic because the models do not accurately reflect student outcomes such as graduation and job attainment. This ultimately suggests that even as these models predict student success, the models are not very effective at determining how students will actually turn out. Rothstein (2010) demonstrated how, impossibly, VAM for the current year can predict teacher performance in previous years. Van de Grift (2009) notes that there were many validity issues with the tests used to measure teacher value-added. An important part of these validity issues centers around random assignment. These models are based on assumptions of normal student distributions which means in order for these tests to be valid students must be assigned to teacher classrooms randomly (Baker et al., 2010). If random assignment is not maintained, then the student distributions become skewed which ultimately invalidates the models. The problem with this requirement is that random assignment is hardly ever utilized. MET project schools might be the only exception (Bill and Melinda Gates Foundation, 2012). Random assignment of students is rarely utilized in schools for a variety of reasons. Often principals reward and punish teachers by purposely assigning academically strong or weak students to specific classrooms. Additionally, parents request specific teachers based on personal assumptions and desires (Jacob & Lefgren, 2007). Also, schooling is usually districted by neighborhood, and neighborhoods are generally
socioeconomically stratified. Furthermore, research suggests that each test must be validated for
the use in its respective VAM (American Educational Research Association, 2000; Schmitz &
Raymond, 2008; Spring, 2008; Whiteman, Shi, & Plucker, 2011).

Unreliable

Reliability is an issue separate from validity. Reliability addresses the issue of
consistency; are the results constant as the test or measurement is repeated? In a research paper
by Kane and Staiger (2008), the prediction power of VAM was shown to be problematic because
teacher effects on student performance diminish over time and make VAM reliable only in the
short run. This is concerning because, as addressed previously, VAM are used to predict student
performance over the long-run. What this ultimately suggests is that yes student performance
can be predicted long-term, but those predictions will oscillate as the measures are repeated. van
de Grift (2009) later confirmed the reliability issues with the tests used to measure teacher value-
added.

Data Issues

In research by Misco (2008), it was shown that VAM were problematic because there
were gaps in the data utilized by the models. These gaps result from a variety of student issues
such as transiency, repeating grades, and homelessness. These issues are social issues that likely
cannot be addressed by teachers. Thus, it is problematic to hold teachers accountable for criteria
they cannot control. This directly fails the first requirement of work conducted by Peterson
(2000) suggesting that there are six tests to determine if data sources are acceptable for VAM:

1. Are the data caused by (or the responsibility of) the teacher?

2. Are the data included in the job description of the teacher?

3. Are the data linked to student learning, welfare, or other needs?
4. Are the data of primary importance in consideration of teacher quality?

5. Do data predict or consistently associate with questions of primary importance?

6. Are better data available on the same issues?

**Floor and Ceiling Effects**

Another factor that corresponds to the validity of these measurements is floor and ceiling effects. Research conducted by Koedel and Betts (2008) showed that VAM have ceiling effects which limit a teacher’s level of effectiveness. A ceiling effect is when a test has an upper limit. For example, if a student takes a math exam and the highest level it tests is eighth grade standards, then the highest score a student can achieve is mastery of eighth grade standards. This is concerning because a teacher could have raised a student’s knowledge beyond that level, but this growth cannot be identified through testing instruments with test ceilings. Additionally, this is dubious for students at the upper-end of the achievement distribution because a teacher cannot show growth for a student who has already achieved mastery of the content. Floor effects are the same but at the lower end of the spectrum. Just as it is important for a test to allow a student to demonstrate how much academic growth they have made, it is equally important for a student to show how much academic regression or non-growth a student had made. If a student is placed in a seventh grade classroom but only has third grade mathematics skills, a floor effect will depict the student as having seventh grade math skills. Thus, if a teacher raised the student from the third grade level to the fifth grade level, a floor effect would yield a result of no growth by placing the student at the seventh grade level.

**Lack of Transparency**

Another major criticism of VAM is that the models utilized are largely proprietary, meaning that the models are not peer-reviewed (Kupermintz, 2003). The inability to analyze the
algorithms behind the models casts doubt upon the beneficial claims made by the model’s creators. Even more troubling, is that a moral hazard could exist. A moral hazard occurs when an individual changes his predetermined behavior through a contractual agreement after the contract is initiated (Michel-Kerjan & Slovic, 2010). In this setting, a model creator could make changes to models to ensure certain outcomes that have associated financial impacts. Without public oversight, there is no way to determine if moral hazards or conflicts of interest exist.

Misidentifying Effective Teachers

The main issue with using value-added teacher evaluations is misidentification of the strongest and weakest teachers. Teachers can be misclassified as effective or ineffective due to inaccuracies in VAM (Guarino, Reckase, & Woolridge, 2012). In an Economic Policy Institute brief on using VAM to make personnel decisions it was stated that:

there is no current evidence to indicate either that the departing teachers would actually be the weakest teachers, or that the departing teachers will improve student learning if teachers are evaluated based on test score gains or are monetarily rewarded for raising scores. (Baker et al., 2010, p. 5)

The reason this is likely the case is that that students’ scores are a result of a wide variety of variables as stated by Baker and others (2010):

Student test score gains are also strongly influenced by school attendance and a variety of out-of-school learning experiences at home, with peers, at museums and libraries, in summer programs, on-line, and in the community. (p. 3)

This idea is also contained in CEM (Boardman & Murnane, 1979; Guarino, Reckase, & Woolridge, 2012; Hanushek, 1979, 1986; Todd & Wolpin, 2003). One important factor is socioeconomic status. Student test scores are influenced by economic issues such as parental
income levels. Thus, teachers teaching students with a higher socioeconomic status are more likely to be rated as more effective than teachers teaching low socioeconomic status children (Baker et al., 2010). Hence, roughly fifty percent of teachers change rankings by more than one quintile year-to-year (Goldhaber & Hansen, 2008). In a study by Mathematica, error rates for teacher misplacement were 36% when one year of data was used to measure the teacher (Schochet & Chiang, 2010). This means that roughly one out of three teachers is categorized as an effective teacher when he is ineffective and vice versa. Additionally, the study’s authors calculated that in order to reduce the error rate to 12%, ten years of value-added data would be necessary (Schochet & Chiang, 2010). Needing ten years of accumulated data to get the best measure of teacher effectiveness is inefficient at best and inappropriate for teacher appraisal at worst. Additionally, an error of 12%, misplacing approximately one out of eight teachers, is unacceptable by any standard of measure. Finally, Raudenbush (2004, p. 124) reported that “it does not appear possible to separate teacher and school effects using currently available accountability data.” This means that school inputs such as class size, curriculum, and programming cannot be separated from the effect teachers have on student achievement.

**Claimed Usage in other Disciplines**

Many claims have been made that VAM should be used in education as most other disciplines outside of education use this approach to evaluate employees. This simply is not true. In fact, management experts have cautioned against the use of quantitative analysis to determine merit pay (Rothstein, Jacobsen & Wilder, 2008). In a study by Bowman (2010), it was shown that when these forms of evaluation are used as they have been via the federal government, they fail miserably to meet their objective of rewarding exceptionalism. Performance pay plans have been paradoxically used for the past 20 years to increase productivity. Unfortunately, the plans
as alluded to earlier have decreased employee morale and led to decreased productivity
(Bowman, 2010).

**Perverse Incentives**

Another problem that arises from the use of high-stakes testing in these models is the
incentive for students to take less rigorous courses and engage in cheating as both students and
teachers receive rewards for high test scores (Nichols & Berliner, 2007). This has the potential
to create an education system structured against the best interest of society which is to develop
and stimulate critical thinking across a multitude of disciplines. If teachers are penalized for
teaching the most troubled students by receiving lowered evaluations due to poor student
achievement, then they will be unwilling to teach the neediest students (Baker et al., 2010).
Again, this acts in direct opposition to the mission of schools to educate all students, not just the
ones that will likely score well on high-stakes standardized tests.

**Not Utilized to Improve Instruction.**

As mentioned previously, VAM have been shown to improve student outcomes.
However, there is little evidence to suggest that teachers and schools are using the data generated
to direct and improve instruction (Raudenbush, 2004). This likely has to do with the complexity
of the models and educators’ ability to understand the meaning of the resultant data (Callendar,
2004). For example, hierarchal high degree polynomial models have been suggested for use in
measuring student achievement and teacher effectiveness (Raudenbush & Bryk, 2002). It is safe
to say that most educators would have apprehension in dealing with the mathematical skill
required to understand, analyze and use hierarchical high degree polynomial models (Battista,
1986). An example is illustrated below:

\[ Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij}) + \beta_{2j}(X_{2ij}) + e_{ij} \]
\[ \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + u_{1j} \]

Because of the model’s mathematical complexity, it is likely that educators and policy makers would ignore or abandon its use regardless of the merit of the model.

**Teaching to the Test**

The probable result of such an increasing focus on student test scores with significant punishments as severe as job loss is that teachers will teach to the test. This means that they will focus solely on test preparation and repetition at the expense of higher-order learning. This could be the teacher’s choice, but what is more problematic is that it is often forced by administrators. In one study it was suggested that school administrators were encouraging teaching to the test to avoid the negative consequences that these tests bring upon schools (Nichols & Berliner, 2007). This is problematic for a variety of reasons, but the main issue is that if administrators are forcing a given instructional method, then VAM are measuring administrators instead of teachers. This again invalidates the use of these models for teacher evaluation purposes. While teaching to the test can be positive in terms of reducing testing anxiety due to test set-up, it likely comes at the expense of higher order thinking (Darling-Hammond & Rustique-Forrester, 2005).

A narrowed curriculum results from teaching to the test. A narrowed curriculum is one that focuses on specific content instead of a well-rounded education. For example, high-stakes tests are usually in math and English. Thus, a narrowed curriculum begins to focus with much more intensity and duration on math and English at the expense of other disciplines such as science and history. Baker and colleagues (2010) reported that this exclusive focus on math and reading tests has caused a narrowing of the curriculum. This will likely have negative
implications for students and society as citizens will be less well-rounded than previous generations. This could have many unintended consequences economically and politically.

**Decreased Motivation**

While VAM might be an improved way to measure teacher effectiveness, their usage may come at a high cost in terms of teacher morale. As Baker and colleagues (2010) pointed out, the use of VAM decreased student and teacher motivation. Additionally, in an article by Anderman, Anderman, Yough, and Gimbert (2010) it was shown that educators are not more motivated through the use of VAM to dispense monetary rewards. These articles along with Tony Bastick’s (2000) work on why teachers teach imply that teachers teach out of an altruist belief that they are improving society. Reducing them to a metric and tying their pay to that measure devalues their contribution to society. This likely will have lasting consequences with decreased student learning and increased teacher attrition.

**Teacher Demoralization**

When all of these negative consequences are viewed collectively, it bears the question will teachers stay in a profession with such poor working conditions. It is likely teachers will become dissatisfied with their working conditions and teachers will leave the profession (Baker et al., 2010). Again, this is an unintended consequence that may have significantly negative effects. Research has shown that humiliating teachers by mandating parental notification of ineffective teaching will damage the teaching profession (Darling-Hammond, Wise, & Pease, 1983; National Council for Teacher Quality, 2011; Whiteman, Shi, & Plucker, 2011). Additionally, a MetLife teacher survey (2013) indicated that teachers’ job satisfaction has declined to an all-time low with a rapid decline from 62% in 2008 to 39% in 2012 saying that they are very satisfied with their job. Again, this survey coincided with RTT implementation
that required VAM to be tied to teacher evaluation in order for states to be eligible for funding.

**Lack of Necessity**

The final point when considering all the negative aspects of value-added measures is how necessary is VAM in education? Previous research has shown there are alternatives. Traditional measures such as teacher experience have shown to be an estimate of teacher effectiveness as there is a positive association between teacher experience and student achievement (Clotfelter, Ladd, & Vigdor, 2006; 2007). Another important factor is pedagogical content knowledge which means that a teacher understands which teaching practices most appropriately fit the particular content being taught and how the content being taught should be ordered to enhance student learning. This has also been shown to be a predictor of student achievement (Hill, Rowan, & Ball, 2005).

National Board Certification is a process in which educators progress through rigorous criteria to demonstrate they are master teachers. This also has been shown to be a valuable measure of teacher effectiveness with National Board Certified teachers being more effective than non-National Board Certified teachers (Goldhaber & Anthony, 2007). Finally, teacher qualifications such as degree attainment have been shown to increase student achievement (Boyd et al., 2008; Croninger, Rice, Rathbun, & Nishio, 2007). Collectively, these traditional measures of teacher quality all have credibility in terms of determining teacher effectiveness without all of the associated negative consequences of value-added teacher evaluations.

**Indiana’s Position on Value-Added Measures**

The issues addressed in the previous sections represent an overall review of the literature on VAM. This has demonstrated that there are many pros and cons with regard to the usage of VAM to evaluate teachers. However, Indiana’s use of VAM is much more focused in that the
state’s model mandates the use of student growth measures to measure teacher performance. The following sections will review Indiana’s version of VAM with respect to the issues identified previously.

**Model Creator Claims**

The switch to a value-added teacher evaluation system in Indiana was proposed and supported by former Governor Mitch Daniels who served from 2005-2013 and former State Superintendent of Education Dr. Tony Bennett who served from 2008-2012 in order to achieve three positive educational outcomes:

1. Evaluate and pay teachers based on student learning.
2. Hold schools accountable for student learning while giving them the flexibility to deliver better results under local control.

While these suggestions have not been validated by research, other factors of Indiana’s VAM model have been studied. In the following sections, the literature focusing on Indiana’s VAM model and growth measures will be reviewed.

**Desired by the Public**

Similar to the Phi Delta Kappa/Gallup Poll referenced previously (Rose & Gallup, 2007), in a survey by Indiana University’s Center for Evaluation and Education Policy (CEEP), 159 out of 179 (88.8%) Indiana superintendents strongly agreed, agreed, or somewhat agreed that teacher evaluation should be linked to student growth (Cole, Robinson, Ansaldo, Whiteman, & Spradlin, 2012). However, it is unknown if Indiana teachers feel the same way. Since, teachers are the individuals responsible for teaching students, their opinions are probably much more important
as suggested by Elmore (2004).

**Invalid**

In order to level the playing field the tests on which growth models are based must be valid and reliable. Under Indiana’s system, they likely will meet neither criterion. Validity is incredibly problematic for Indiana educators as only math, English, social studies, and science have statewide assessments. This means that all other disciplines will require district, school, or teacher created tests to be used. These tests will likely not meet the standards of validation to be used because test-validity is difficult to achieve even for experts. According to Messick (1995), the two main sources of teacher-created test invalidity are construct under-representation (CUR) and construct-irrelevant variance (CIV). CUR is defined as a test not having enough items dealing with the intended criterion to be measured on the test to measure the criterion. An example of this could be a test that only has low-level questions when it is intended to measure higher-order thinking. CIV is defined as a test that requires skills that it does not intend to measure. An example of this would be a history test that requires reading passages for a student that has vision problems. If the student cannot read the test, his ability to successfully answer the questions cannot be measured. Requiring teachers who have many additional daily tasks to create tests that measure up to these validity standards is not a sound strategy. In fact, in a WTHR Indiana teacher poll 2,824 out of 4,278 (66.0%) said they do not think it is fair to hold teachers accountable for their students’ academic achievement (WTHR, 2011). While there is no indication that this poll was representative of all Indiana teachers, it does indicate a potential issue teachers may be concerned about.

**Increased Cost Efficiency**

Increased cost efficiency is unlikely in Indiana’s case. The requirements of the new
evaluation plan are substantial as indicated in the Lack of Transparency subsection of the Criticisms of Value-Added Measures section of this paper. These requirements will likely require a large capital outlay and continuance that will probably decrease cost efficiency as tests, data management systems, additional staff, and increased evaluation model task performance will have to be expensed.

**Data Issues**

The authors of a CEEP policy brief suggest that under Indiana’s model, likely none of Peterson’s (2000) criteria are met for acceptable data sources of VAM (Whiteman, Shi, & Plucker, 2011):

1. The standardized test results are not solely caused by the teacher as referenced previously.
2. The standardized test results are not likely to be included in the job description of the teacher.
3. The standardized test results are not likely linked to student learning, welfare, and other needs.
4. The standardized test results are not of primary importance in consideration of teacher quality as referenced previously.
5. The standardized test results do not predict or consistently associate with questions of primary importance as referenced previously.
6. Better data is available to predict teacher quality as referenced previously.

**Mandated Failure**

Indiana’s model, at the time of this writing, focused on student growth from test to test. Student growth was ordered from least to greatest, and then the median score was set as the
benchmark. This methodology mandates that at least half of the students will not meet the expected growth criterion. This is problematic because, “If it is assumed that all schools strive to improve the academic growth of their students, then using median improvement as the criterion for evaluating change masks the actual progress made by many students and schools” (Anderman, Anderman, Yough, & Gimbert, 2010, p. 135). Ultimately, this new method has the potential to shift the educational paradigm in a negative way. It will likely switch the paradigm to an educational focus of improving the mid-level students while excluding the top and bottom because their scores are not relevant in determining teacher effectiveness.

**Lack of Transparency**

Another major criticism of VAM is that largely the models utilized are proprietary, meaning that the models are not able to be peer-reviewed (Kupermintz, 2003). The inability to analyze the algorithms behind the models casts doubt upon the beneficial claims made by the model’s creators. In Indiana’s case, the lack of transparency is derived from the State Board of Education’s (SBOE) role in defining and establishing:

- Criteria defining each of the teacher ratings (Highly Effective, Effective, Improvement Needed, Ineffective),
- Measures used to determine academic growth,
- Standards defining a teacher’s negative impact (Whiteman, Shi, & Plucker, 2011).

The SBOE can redefine these areas year-to-year to acquiesce to political whims as they see fit. Thus, year-to-year, teachers and administrators cannot determine what metric their students must meet to be placed into categories that will determine pay and continued employment.

**Less Teacher Collaboration**

A consequence of value-added teacher evaluations is that teachers are forced to compete
for limited resources in terms of annual raises. Teachers who are rated highly will be given the biggest raises. In fact, in Indiana, teachers rated effective or highly effective will receive a pay increase from the pool of teachers rated needs improvement or ineffective (Indiana Department of Education, 2012). Thus, there is the potential for any teacher looking to maximize his yearly raise to have a vested interest in his colleagues performing poorly. Intuitively, this seems like a strong idea. The best teachers should be rewarded with the best pay, but teaching is collaborative with teachers of different disciplines working together to educate the whole student. As these measures are utilized to determine teacher pay, the career could become more competitive. Thus, it would be counter-productive for one teacher to continue to help another teacher because he or she would be acting against his or her own economic self-interest. This was reported in a policy brief that asked why teachers would help each other when they are being compared to one another (Baker et al., 2010). However, it is unknown if Indiana teachers feel this way and are engaging in such practices.

**Unwillingness to Accept Student Teachers**

Corresponding to less teacher collaboration, there is also increasing fear among schools of education that as the Indiana model is adopted and implemented by schools, teachers will become increasingly unwilling to take on student teachers due to the concern that the student teacher could negatively affect student outcomes (Whiteman, Shi, & Plucker, 2011). This could be one of the biggest issues the new model faces as student teaching is a teacher licensing requirement in the state of Indiana (Indiana Code 20-28-2-6).

**Stunted Student Growth**

During the teacher evaluation model implementation year, schools are required to meet a multitude of requirements:
• Establish common expectations for teaching standards and methods of evaluating teachers’ attainment of those standards;

• Train an evaluation team and develop methods of ensuring inter-rater reliability;

• Determine frequency and standards for teacher observations, methods and format for teacher feedback, and a due process procedure;

• Define student growth;

• Develop valid and reliable local assessments for all coursework for the purpose of determining student growth;

• Build or purchase information management infrastructure to collect teacher evaluation data and student performance data;

• Train teachers, counselors, principals, and the evaluation team to access, interpret, and appropriately act on teacher evaluation and student performance data;

• Develop a system for synthesizing teacher activities and instructional practices (inputs) with student growth and instructional outcomes (outputs) to form a summative rating of Highly Effective, Effective, Needs Improvement, or Ineffective;

• Establish procedures for distributing compensation based on summative classifications (Whiteman, Shi, & Plucker, 2011).

Meeting all of these requirements will be incredibly difficult for schools to accomplish considering administrator and teacher psychometric inability (Smith, 2002). Thus, it is highly likely that during the implementation year and thereafter, student learning will be negatively impacted because of the educational time requirement shift from student achievement to model implementation.
Legal Issues

In Indiana VAM will, at least, partially be used to make employment decisions about teachers (Indiana Code 20-28-7.5-1). Inevitably, some teachers will lose their jobs due to poor performance evaluations resulting from VAM. Potentially, when an employee feels he has been wrongfully terminated, the employee sues. The situation in Indiana could result in an onslaught of litigation for a variety of reasons involving violations of Federal and Indiana law. Every issue noted in the previous section, individually, could be used as evidence of wrongful termination. Collectively, and with expert testimony, it would be difficult to counter the arguments made against VAM. The following sections will detail which laws will be cited in bringing suits against schools and administrators.

Title VII. Title VII of the 1964 Civil Rights Act establishes that, “employment decisions cannot have a disparate impact on protected groups” such as those comprised of specific groups of minority teachers (United States, 1968). If data could be accumulated, and it probably can considering all the negative research of VAM that currently exists, to show that any of these groups have consistently lower VAM scores, this could be a strong legal argument. There is a great deal of evidence that suggests that disadvantaged groups are segregated in schools (e.g. Kozol, 1992; Orfield & Lee 2005). If that research could be coupled with research showing that the segregated groups of students routinely underperform on the specific assessments used to evaluate teachers, this could be extremely damaging evidence against the use of VAM to evaluate teachers. However, the differences in VAM across racial groups would not be enough to make the use of VAM-based teacher evaluations illegal. According to one prominent education lawyer Scott Baueries (2010), “if an employer can show that the VAM is job-related, meaning that it fairly and accurately measures a teacher’s ability to perform the essential
functions of his or her job, and is consistent with business necessity then unequal outcomes would pass muster under Title VII.” In Indiana’s case, a strong argument could also be made that the evaluation model does not accurately measure a teacher’s ability to perform the essential functions of his or her job which are defined by the model as an ability to increase median student learning growth as measured by authorized assessments as indicated in the Invalid subsection of this paper.

**Due process.** Another area of litigation focus with regard to VAM-based teacher evaluations is due process which is written into the Fifth and Fourteenth Amendments of the United States Constitution. Both clauses prohibit the Federal and state governments from depriving any person of "life, liberty, or property, without due process of law” (United States, 2006). Due process falls into two categories procedural due process (PDP) and substantive due process (SDP). Both areas are possible areas of litigation with respect to teacher evaluations. However, SDP is more likely to be used as an argument against the Indiana VAM-based teacher evaluations.

Procedural due process provisions require that the same procedure for employment decisions be used for all potential applicants. School districts must show that they have provided employees with sufficient notice of impending termination of demotion and an opportunity to be heard usually in a hearing before the School Board or it’s designate. Indiana Code specifically addresses this:

A copy of the completed evaluation, including any documentation related to the evaluation, must be provided to a certificated employee not later than seven (7) days after the evaluation is conducted. If a certificated employee receives a rating of ineffective or improvement necessary, the evaluator and the certificated employee shall develop a
remediation plan of not more than ninety (90) school days in length to correct the deficiencies noted in the certificated employee's evaluation. The remediation plan must require the use of the certificated employee's license renewal credits in professional development activities intended to help the certificated employee achieve an effective rating on the next performance evaluation. If the principal did not conduct the performance evaluation, the principal may direct the use of the certificated employee's license renewal credits under this subsection. A teacher who receives a rating of ineffective may file a request for a private conference with the superintendent or the superintendent's designee not later than five (5) days after receiving notice that the teacher received a rating of ineffective. The teacher is entitled to a private conference with the superintendent or superintendent's designee. (Indiana Code 20-28-11.5)

VAM would not directly impact a principal’s ability to meet these requirements. However, this new evaluation model requires principals to observe teachers much more frequently than in the past. With an increased workload (Indiana Department of Education, 2012a), it may be difficult for some administrators to meet the requirements of sufficient notice and a hearing in the dictated time frames. This type of defense would not be useful on a large scale because it does not demonstrate fault in the evaluation method, but individual fault by specific administrators.

Substantive due process prevents decisions from being “arbitrary or capricious deprivation of property,” (United States, 2006) which is often defined to include employment. A strong case based on arbitrariness could be built based on the makeup and results of VAM’s random error and assessments content. If the assessments and results used to place teachers in categories that dictate compensation and continued employment could be shown to be arbitrary, then litigants would likely be victorious in wrongful termination cases based on Indiana VAM-
based teaching evaluations. As noted in the Criticisms of VAM section of this paper, it is obvious that there are serious psychometric problems with VAM and most judges with a well-reasoned argument would rule that the VAM part of the evaluation is in fact arbitrary. Additionally, since Indiana Code requires, “A provision that a teacher who negatively affects student achievement and growth cannot receive a rating of highly effective or effective” (Indiana Code 20-28-11.5), and, as noted previously, teacher pay and retention being based on the effectiveness rating, pay and retention would logically be deduced as being arbitrary.

Teacher Demoralization

When all of these negative consequences are viewed collectively, it again illustrates the likelihood that some teachers will not want to face these issues and teachers will leave the profession (Baker et al., 2010). Again, this is an unintended consequence that may have significantly negative effects. Research has shown that humiliating teachers by mandating parental notification of ineffective teaching will damage the teaching profession (Darling-Hammond, Wise, & Pease, 1983; National Council for Teacher Quality, 2011; Whiteman, Shi, & Plucker, 2011). In fact, in a WTHR Indiana teacher poll (2011) 1,375 out of 4,363 (31.5%) said they would not choose teaching if they could do it all over again and 795 out of 4,297 (18.5%) reported they plan to change careers. While participation in this poll was self-selected and thus not representative, it still illustrates some degree of teacher dissatisfaction with the state of the teaching profession.

Summary

As indicated through this literature review, published research on VAM has taken on a largely quantitative and critical point of view. Most published research has dealt with the statistical limitations of VAM in that these measures provide results that are generally neither
valid nor reliable. Additionally, many articles have demonstrated problems resulting from sampling, test construction, and stretch. A test that does not have sufficient stretch has ceiling and floor effects (Koedel & Betts, 2008). This means that achievement is constrained by test question difficulty. If a test has a ceiling or floor effect then true student educational attainment is constrained, thus limiting the impact a teacher can have on student growth. All of these issues are cause for concern. CEEP released a policy brief that investigated school administrator perceptions about the new evaluation method being used in Indiana (Cole, Robinson, Ansaldo, Whiteman, & Spradlin, 2012). This study offers a wealth of information as to how administrators view the model, but it does not include teacher perceptions at all.

Research on Indiana’s VAM system has not been conducted on two key areas. First, no research has been done on VAM to determine how these models, used within a teaching evaluation framework, impact teachers. However, many anecdotes have been submitted in periodicals and online forums on the Internet suggesting that teachers are dissatisfied with this form of evaluation. No formal research studies have been undertaken in an effort to understand how teachers are impacted by value-added teacher evaluations and what results these teacher impacts have on student education in general.

Also, no research has been undertaken on value-added teacher evaluations in Indiana to investigate the impacts on student academic achievement to better understand whether or not VAM is improving educational outcomes in Indiana.

**Relevance of Teacher Perceptions**

According to the Joint Committee on Standards for Educational Evaluation (1988) an evaluation system must have four attributes: propriety, utility, feasibility, and accuracy.
Propriety ensures that teachers are treated respectfully and conflicts of interest are managed under teacher performance review. Utility allows teacher growth and improvement instead of simply applying a summative rating. Feasibility yields necessary data with minimum disruption to instructional processes at the lowest practical costs. Accuracy minimizes bias brought by the system’s mechanics and/or the evaluator (Whiteman, Shi, & Plucker, 2011).

As illustrated in this literature review, Indiana’s VAM as part of a teacher evaluation framework fails on all four domains. It fails the propriety requirement as teachers are not treated respectfully. They are treated as replaceable inputs in a production function also referred to as the Widget Effect (Weisberg, Sexton, Mulhern, & Keeling, 2009). This is being done by holding teachers solely responsible for student achievement results. It fails the utility requirement by effectively applying a summative rating if student growth or assessment does not meet the required level by mandating that a teacher not receive a pay increase or have their employment terminated. The model fails the feasibility requirement as students are forced to take a multitude of tests in all subjects which will considerably narrow the curriculum. Finally, it fails the accuracy standard as VAM are both invalid and unreliable.

This leaves a fundamental question left unanswered as to how Indiana teachers perceive Indiana’s VAM-based model and the use of growth model data in teaching evaluations. This is important for many reasons. In terms of the CEM variant in which teachers are viewed as the single variable of the education production function, if the model impacts teacher’s perceptions in a negative way, then the production created by teachers will be diminished. If the model improves teachers’ self-worth and job satisfaction, then production will be increased. Thus, it is incredibly important to understand how teachers in Indiana feel about the use of the VAM-based model. This ultimately leads to the second fundamental question which is probably more
important. Does the use of a VAM-based teacher evaluation increase student achievement? Ultimately, educational practices should result in increased student achievement. If that does not occur then alternate methods should be utilized.

Relevance of Increased Student Achievement

While it has been suggested that VAM has the potential to increase student achievement, these studies have focused on VAM as a tool for self-evaluation (Demie, 2003) and that high teacher VAM scores result in increased student achievement for students in that classroom (Aaronson, Barrow, and Sander, 2007; Hanushek & Rivkin, 2008), no studies investigated Indiana VAM’s impact on student achievement as a whole. Increased student achievement is an incredibly important as it is one of the main focuses of an educational system as a whole. If student achievement is consistently raised under this system, then negative externalities may be a price worth paying. However, if student achievement is not consistently raised and there are additional negative externalities, then the model may need to be reevaluated or even eliminated.

Research Questions

Five fundamental unanswered questions have risen from the literature review of the use of growth measures to evaluate teachers under Indiana’s system. The specific research questions addressed in this dissertation are:

1) Is VAMIN resulting in better student achievement when compared to the previous year’s non-VAMIN teacher evaluation?

2) Is VAMIN resulting in better student growth when compared to the previous year’s non-VAMIN teacher evaluation?

3) Are there relationships between student achievement growth under a VAMIN-based teacher evaluation when controlling for student gender, race, special education status,
socioeconomic status, Section 504 status, English learner status, and student previous year achievement score?

4) Is there consistency between teacher effectiveness ratings and principal-based observation evaluations?

5) What impacts do teachers perceive that VAMIN has on teachers in VAMSD?

**Conclusion**

This literature review focused on the analysis of past and current research dealing with VAM particularly in relationship to teacher evaluations. After analyzing relevant research, it is now obvious that there is an important gap in published research about the effects of value-added teacher evaluations on school teachers, and that such a project is a worthwhile dissertation topic. This review also identified many potential teacher perspectives that will likely be addressed during future research endeavors ranging from reliability and validity issues to missing data.

The major contributions of previous research ultimately question the appropriateness of using value-added measures to evaluate teachers (Baker et al., 2010), review the laws surrounding the change in Indiana’s teacher evaluation model (Whiteman, Shi, & Plucker, 2011), and detail administrator perceptions about the use and implementation of the models (Cole, Robinson, Ansaldo, Whiteman, & Spradlin, 2012). Additionally, this will likely lead to continued research in this area as a number of working papers addressing value-added measures in education are increasingly undertaken.

After thorough analysis, it is apparent that research on teacher perceptions about the use of VAM to evaluate teachers is lacking. As these measures increase in implementation and the increasing questionability of their appropriateness becomes more apparent, it will be imperative
for researchers to determine what effects this evaluation method is having on teachers, both benefits and detriments.

Ultimately, through this literature review, it has been demonstrated that future research endeavors dealing with teacher perceptions about using value-added teaching evaluations is imperative. This literature review represents the first step in spurring additional research focused on the possibility of identifying the hopes and fears that teachers have about the use of VAM as part of evaluation models. Hopefully, this future research can give voice to authentic teacher perceptions that could be used to improve VAM for evaluation purposes. Additionally, this research has the potential to spur additional research which could potentially shape future educational policy and provide insight on how the conversation about implementing value-added teaching evaluations should be implemented to increase teacher buy-in.
CHAPTER 3

RESEARCH METHODS

Most of the research on VAM has focused on the mathematical nature of the problem (Baker et. al., 2010; Broatch & Lohr, 2012; Kane & Staiger, 2008; Koedel & Betts, 2008; Konstantopoulos & Sun, 2012; Lefgren & Sims, 2012; van de Grift, 2009; Winters, Dixon & Greene, 2012). No peer-reviewed research was found that investigated teacher’ perceptions about the use of VAM to evaluate teachers, and no research was found that investigated the impacts of VAM-based teacher evaluations on student academic achievement in Indiana.

The research conducted in this study is focused on investigating the impact of a value-added teacher evaluation model on teachers in a large urban Indianapolis school district. This research was conducted through five distinct investigations:

Is VAMIN resulting in better student achievement when compared to the previous year’s non-VAMIN teacher evaluation? This investigation will be referred to as the VAMIN Evaluation Type Comparative Study (VETCS) for the remainder of this study.

Is VAMIN resulting in better student growth when compared to the previous year’s non-VAMIN teacher evaluation? This investigation will be referred to as the VAMIN Growth Comparative Study (VGCS) for the remainder of this dissertation.

Are there relationships between student achievement growth under a VAMIN-based teacher evaluation when controlling for student gender, race, special education status,
socioeconomic status, Section 504 status, English learner status, and student previous year achievement score? This investigation will be referred to as the VAMIN Multiple Regression Study (VMRS) for the remainder of this study.

Is there consistency between teacher effectiveness ratings and principal-based observation evaluations? This investigation will be referred to as the VAMIN Consistency Correlational Study (VCCS) for the remainder of this dissertation.

What impacts do teachers perceive that VAMIN has on teachers in VAMSD? This investigation will be referred to as the VAMIN Survey of Teacher Impacts (VSTI) for the remainder of this study.

This chapter includes the following sections: sample, variables, measures, statistical relationships, and limitations for each investigation. The sample section details what types of samples or participants were used, the sampling design, and power analyses. The variables section details the variables used in the investigation. The measures section details the statistical measures utilized with applicable procedures. The statistical relationships section discusses the analytic method section for each investigation. The limitations section details generalizability restrictions and aspects over which the researcher had no control.

VETCS

Sample

The samples used to conduct the investigation were comprised of approximately 5,000 VAMSD student test scores on the English language arts (ELA), math, science, and social studies Indiana Statewide Testing for Educational Progress Plus (ISTEP+) tests in grades three through eight during the 2011-2012 and 2012-2013 school years. The grouping of the samples used to conduct the analysis is detailed below in Table 3.1. These comparisons were cohort
comparisons of subsequent years, third grade compared to third grade. The samples were essentially census samples in that they were comprised of all grade-level student ISTEP+ scores used under the VAMIN method in 2012-2013 and the non-VAMIN method in 2011-2012. The sample data was provided by the Indiana Department of Education via the Indiana Online Reporting System (INORS).

The groupings for the sample were all approximately normal since each was large (700-plus) and comprised of student testing data. The smallest sample had a size of \( n = 727 \). Thus, even if the samples were not normally distributed, they met the requirement for being able to apply the Central Limit Theorem by having a size of at least \( n = 30 \) (Starnes, Yates, & Moore, 2012). An analysis of sample characteristics was conducted prior to the use of the \( t \)-test. Table 3.1 below details the number of elements included in the groupings of samples used in this analysis.

<table>
<thead>
<tr>
<th>VETCS Sample Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Level</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>
Variables

VAMSD student ISTEP+ test scores was the dependent variable. It had an interval scale and its range of values for each grade and subject tested is detailed in Appendix C. Teacher evaluation method was the independent variable. It had a nominal scale because it had dichotomous labels of VAMIN and non-VAMIN.

Measures

Descriptive univariate statistics were calculated for the ISTEP+ score variable being analyzed in this research to give an idea of the variable’s center, shape, and spread. This information was compared to scatter plot data to ensure approximate normality.

Table 3.2 provides information on the statistical procedures used based on the type of variable compared to the continuous variables of student growth measures. The table details how the comparative variable was treated as a discrete or continuous variable. Additionally, the number of distinct categories the variable maintained is included. This information was combined to determine the specific test statistic and statistical procedures used.

<table>
<thead>
<tr>
<th>VETCS Test Statistic and Statistical Procedure to be Used</th>
<th>Discrete/Continuous</th>
<th>Categories</th>
<th>Test Statistic</th>
<th>Statistical Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Evaluation Method (VAMIN/non-VAMIN)</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>Two Sample $t$-Test</td>
<td>Compare Means $t$-Test Box Plots Spearman Correlation ($\rho$)</td>
</tr>
</tbody>
</table>

Statistical Relationships

The independent samples $t$-test was used to determine the relationship between variables. The mean was used since the relationship between test score and evaluation method was a continuous-to-discrete comparison with two categories. Additionally, the significance level ($p$)
ultimately determined whether the measures were statistically significant or not when measured against an $\alpha$-level of .05. This means that results obtained are believed to have occurred by chance less than five percent of the time.

The independent samples $t$-test is a test of the null hypothesis that the population means related to two independent, random samples from an approximately normal distribution are equal (Altman, 1991).

Power for the independent samples $t$-test was calculated as the power achieved with the given sample sizes and variances for detecting the observed difference between means with a two-sided type I error. This was calculated by taking the complement of the confidence interval (Dupont & Plummer, 1990).

**Limitations**

This analysis is generalizable to teachers in large diverse urban districts who teach students in grades three through eight who take the ISTEP+ exam. The analysis had several potential confounding or lurking variables. However, these should be discounted as this analysis was used to determine if the VAMIN method improved student academic achievement as suggested by the teacher being the only variable in the education production function which is mirrored in Indiana’s use of VAM-based teacher evaluations. So if there were variables that could have confounded the results, these should only be used as a contradiction to the idea that the teacher was solely responsible for educational output.

**VGCS**

This analysis was very similar to the analysis in the VETCS section. However, this analysis focused on student growth that occurred using a VAMIN evaluation method versus a non-VAMIN method.
Sample

The samples used to conduct the investigation were comprised of approximately 3,500 VAMSD growth measures of student test scores on the ELA and math ISTEP+ tests between consecutive grade levels in grades three through eight from the 2010-2011 to the 2011-2012 school year and the 2011-2012 to the 2012-2013 school year. Science and social studies growth measures were not calculated because they were not tested in consecutive grade levels. The grouping of the samples used to conduct the analysis is detailed below in Table 3.3. These comparisons were cohort comparisons of subsequent years (third to fourth grade growth compared to third to fourth grade growth). The samples were essentially census samples in that they were comprised of all grade-level student ISTEP+ growth scores used under the VAMIN method from 2011-2012 to 2012-2013 and the non-VAMIN method from 2010-2011 to 2011-2012. The sample data was provided by the Indiana Department of Education via INORS.

The groupings for the sample were all approximately normal since each was large (600-plus) and comprised of student testing data. The smallest sample had a size of $n = 600$. Thus, even if the samples were not normally distributed, they met the requirement for being able to apply the Central Limit Theorem by having a size of at least $n = 30$ (Starnes, Yates, & Moore, 2012). An analysis of sample characteristics was conducted prior to the use of the $t$-test. Table 3.3 below details the number of elements included in the groupings of samples used in this analysis.
Table 3.3

<table>
<thead>
<tr>
<th>VGCS Sample Grouping</th>
<th>Grade Span</th>
<th>Subject</th>
<th>2011-2012 to 2012-2013</th>
<th>2010-2011 to 2011-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3-4</td>
<td>ELA</td>
<td>658</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td>Math</td>
<td>660</td>
<td>627</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td>ELA</td>
<td>619</td>
<td>687</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td>Math</td>
<td>632</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>5-6</td>
<td>ELA</td>
<td>684</td>
<td>665</td>
</tr>
<tr>
<td></td>
<td>5-6</td>
<td>Math</td>
<td>690</td>
<td>678</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>ELA</td>
<td>698</td>
<td>658</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>Math</td>
<td>675</td>
<td>623</td>
</tr>
<tr>
<td></td>
<td>7-8</td>
<td>ELA</td>
<td>676</td>
<td>619</td>
</tr>
<tr>
<td></td>
<td>7-8</td>
<td>Math</td>
<td>675</td>
<td>623</td>
</tr>
</tbody>
</table>

Variables

VAMSD growth in student ISTEP+ test scores was the dependent variable. It had a ratio scale because it had a true non-arbitrary zero, representing no growth in scores over consecutive years. Teacher evaluation method was the independent variable. It had a nominal scale because it had dichotomous labels of VAMIN and non-VAMIN.

Measures

Descriptive univariate statistics were calculated for the ISTEP+ score variable analyzed in this research to give an idea of the variable’s center, shape, and spread. This information was compared to scatter plot data to ensure approximate normality.

Table 3.4 provides information on the statistical procedures used based on the type of variable compared to the continuous variables of student growth measures. The table details how the comparative variable was treated as a discrete or continuous variable. Additionally, the number of distinct categories the variable maintained is included. This information was combined to determine the specific test statistic and statistical procedures used.
Table 3.4

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discrete/Continuous</th>
<th>Categories</th>
<th>Test Statistic</th>
<th>Statistical Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Evaluation Method (VAMIN/non-VAMIN)</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>Two Sample t-Test</td>
<td>Compare Means t-Test, Box Plots, Spearman Correlation (ρ)</td>
</tr>
</tbody>
</table>

Statistical Relationships

The independent samples $t$-test was used to establish the relationship between variables. The mean was used since the relationship between growth in test scores and evaluation method was a continuous-to-discrete comparison with two categories. Additionally, the significance level ($p$) ultimately determined whether the measures were statistically significant or not when measured against an $\alpha$-level of .05. This means that results obtained are believed to have occurred by chance less than five percent of the time.

The independent samples $t$-test is a test of the null hypothesis that the population means related to two independent, random samples from an approximately normal distribution are equal (Altman, 1991).

Power for the independent samples $t$-test was calculated as the power achieved with the given sample sizes and variances for detecting the observed difference between means with a two-sided type I error. This was calculated by taking the complement of the confidence interval (Dupont & Plummer, 1990).

Limitations

This analysis is generalizable to teachers in large diverse urban districts who teach students in grades three through eight who take the ISTEP+ exam. The analysis did have several potential confounding and lurking variables. However, these should be discounted as this
analysis was used to determine if the VAMIN method improved growth in student academic achievement as suggested by the teacher being the only variable in the education production function. So if there were variables that could confound the results, these should only be used as a contradiction to the idea that the teacher is solely responsible for educational output.

**VMRS**

**Sample**

The samples used to conduct the investigation were the same samples used in VGCS analysis. Additionally, gender, race, special education status, free and reduced lunch status, Section 504 status, English learner status, and the student’s previous year ISTEP+ score were matched to each student test score under VAMIN and non-VAMIN. Samples, again, were comprised of approximately 3,500 VAMSD growth measures of student test scores on the English language arts and math ISTEP+ tests between consecutive grade levels in grades three through eight from the 2010-2011 to the 2011-2012 school year and the 2011-2012 to the 2012-2013 school year. Science and social studies growth measures were not calculated because they were not tested in consecutive grade levels. The grouping of the samples used to conduct the analysis is detailed above in Table 3.3. These comparisons were cohort comparisons of subsequent years, third to fourth grade growth compared to third to fourth grade growth. The samples were essentially census samples in that they were comprised of all grade-level student ISTEP+ growth scores used under the VAMIN method from 2011-2012 to 2012-2013 and the non-VAMIN method from 2010-2011 to 2011-2012. The sample data was provided by the Indiana Department of Education via INORS.

The groupings for the sample were approximately normal since each was large (600-plus) and comprised of student testing data. The smallest sample had a size of $n = 600$. Thus, even if
the samples were not normally distributed, they met the requirement for being able to apply the Central Limit Theorem by having a size of at least \( n = 30 \) (Starnes, Yates, & Moore, 2012). An analysis of sample characteristics was conducted prior to the use of the \( t \)-test. Table 3.4 above details the number of elements included in the groupings of samples used in this analysis. A scatterplot and skewness calculation was done for each sample via the statistical package SPSS to ensure approximate normality.

**Variables**

VAMSD growth in student ISTEP+ test scores was the dependent variable. It has a ratio scale because it has a true non-arbitrary zero, representing no growth in scores over consecutive years. Teacher evaluation method, gender, race, special education status, free and reduced lunch status, Section 504 status, English learner status, and the student’s previous year ISTEP+ score were independent variables. Teacher evaluation method gender, race: white and non-white, special education status, free and reduced lunch status, Section 504 status, and English learner status all have nominal scales because they had dichotomous labels. Student’s previous year ISTEP+ score had an interval scale and its range of values for each grade and subject tested is detailed in Appendix C.

**Measures**

Descriptive univariate statistics were calculated for the variables analyzed in this research to give an idea of the variable’s center, shape, and spread. This information was compared to scatter plot data to ensure approximate normality.

Table 3.5 provides information on the statistical procedures used based on the type of variable compared to the continuous variables of student growth measures. The table details how the comparative variable was treated as a discrete or continuous variable. Additionally, the
number of distinct categories the variable maintains is included. This information was combined to determine the specific test statistic and statistical procedures used.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Discrete/Continuous</th>
<th>Categories</th>
<th>Test Statistic</th>
<th>Statistical Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Evaluation Method (VAMIN/non-VAMIN)</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>z-test</td>
<td>Compare Means z-test</td>
</tr>
<tr>
<td>Gender</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>z-test</td>
<td>Compare Means z-test</td>
</tr>
<tr>
<td>Race (White/non-White)</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>z-test</td>
<td>Compare Means z-test</td>
</tr>
<tr>
<td>Special Education Status</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>z-test</td>
<td>Compare Means z-test</td>
</tr>
<tr>
<td>Free and Reduced Lunch Status</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>z-test</td>
<td>Compare Means z-test</td>
</tr>
<tr>
<td>Section 504 Status</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>z-test</td>
<td>Compare Means z-test</td>
</tr>
<tr>
<td>English Learner Status</td>
<td>Discrete because it uses a nominal scale of measurement</td>
<td>2</td>
<td>z-test</td>
<td>Compare Means z-test</td>
</tr>
<tr>
<td>Previous Year Test Score</td>
<td>Continuous because it uses a ratio scale of measurement</td>
<td>Infinite</td>
<td>t-statistic</td>
<td>Scatter Plot Pearson Correlation (r) Regression</td>
</tr>
</tbody>
</table>
Multiple Regression

There are eight major assumptions of multivariate regression. It should be noted that these assumptions are not uniformly accepted by all statistical researchers as little is in inferential statistics. The first assumption is that the continuous variables must approximate a normal distribution. This is debatable as most statisticians only require the error term of the regression equation to follow a normal distribution. The Central Limit Theorem says that the error term will be normal if the sample is randomly selected and is sufficiently large. In this sample it was safe to assume the sample would be sufficiently large and the samples were census samples so random sampling was unnecessary. The second assumption is that the dependent variable is a linear function of the independent variables or that there is a straight line relationship between the dependent and independent variables. The third assumption is that the error term is unrelated to the independent variables. The fourth assumption of multiple regression is homoscedasticity. This means that the dependent variable has roughly uniform variability for each value of the independent variables. Another way of saying this is that the model residuals will be evenly distributed. The fifth assumption is that disturbances of cases are unrelated. Again this is not a problem since a census was used. A sixth assumption is that the error terms are normally distributed which again was met by the sampling method. A seventh assumption is that there is no collinearity among the independent variables or no independent variables have strong correlations. The final assumption is that there are no extreme outliers. Regression equations are not resistant so extreme outliers can have major impacts on the systems.

The purpose of multivariate regression is to determine the extent of the linear relationship between the dependent variable and independent variables. The equation provides an idea of how one variable relates to another variable in terms of a linear relationship.


**Statistical Relationships**

Multiple regression was used to determine the relationship between variables. The straight-line relationship between variables was used to determine the correlation between variables. The correlation coefficient ($r$) for each variable identifies the relationship between the variables with a range from -1 to 1. The sign indicates the direction of association and the magnitude represents the strength of the relationship. The regression equation with correlation coefficients, and coefficients of determination ($R^2$) was calculated. Additionally, the significance level ($p$) ultimately determined whether the measures were statistically significant or not when measured against an $\alpha$-level of 0.05.

Power for multiple regression is the complement of the probability of making a Type II error (Cohen, 1988). Thus, this calculation was made using SPSS.

**Limitations**

This analysis is generalizable to all teachers who teach in similar districts. The main limitations of using multiple regression are that regression only establishes that there is a relationship not that it is causal and there may be other variables more responsible for the relationship that were untested.

**VCCS**

**Sample**

The samples used to conduct the investigation were comprised of approximately 150 VAMSD summative teacher evaluation ratings in grades four through eight and Teacher Effectiveness Ratings (TER) as calculated by IDOE. The samples were census samples of VAMSD grade four through eight math and ELA teachers teaching at a VAMSD school during the 2012-2013 school year. The sample data was provided by the Indiana Department of
Education via Licensing Verification and Information System (LVIS) and the VAMSD Human Resources Department. Correlation analysis was conducted to determine the strength and direction of the relationship between evaluation components.

Again, these samples were approximately normal since each was sufficiently large, greater than 10, and were comprised of teacher rating data. An analysis of sample characteristics was conducted prior to calculating correlation coefficients. A scatterplot and skewness calculation was performed for each sample via the statistical package SPSS to determine if approximate normality existed.

**Variables**

VAMSD teacher evaluation summative scores and TER scores were the variables for this investigation. Both had interval scales because they had values ranging from one to four.

**Measures**

Descriptive univariate statistics were calculated for the variables analyzed in this research to give an idea of the variable’s center, shape, and spread. This information was compared to scatter plot data to determine approximate normality.

Table 3.6 provides information on the statistical procedures used based on the type of variable compared to the continuous variables of student growth measures. The table details how the comparative variable was treated as a discrete or continuous variable. Additionally, the number of distinct categories the variable maintained was included. This information was combined to determine the specific test statistic and statistical procedures used.
Table 3.6

VCCS Test Statistic and Statistical Procedure to be Used

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discrete/Continuous Categories</th>
<th>Test Statistic</th>
<th>Statistical Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAMSD Teacher Evaluation Summative Scores from 2012-2013</td>
<td>Continuous because it uses an interval scale of measurement</td>
<td>Infinite</td>
<td>t-Statistic</td>
</tr>
<tr>
<td>VAMSD Teacher Evaluation TER Scores from 2012-2013</td>
<td>Continuous because it uses an interval scale of measurement</td>
<td>Infinite</td>
<td>t-Statistic</td>
</tr>
</tbody>
</table>

**Statistical Relationships**

The straight-line relationship between variables was used to determine the correlation between variables. The correlation coefficient \((r)\) for each variable identifies the relationship between the variables with a range from -1 to 1. The sign indicates the direction of association and the magnitude represents the strength of the relationship. Correlation coefficients and coefficients of determination \((R^2)\) were calculated. Additionally, the significance level \((p)\) ultimately determined whether the measures were statistically significant or not when measured against an \(\alpha\)-level of .05.

**Limitations**

This analysis is generalizable to teachers teaching in large diverse urban metropolitan school districts. The main limitations of using Pearson Correlation Coefficients is that correlation only establishes that there is a relationship not that it is causal and there may be other variables more responsible for the relationship that were untested.
VSTI

Strategy

The participants in this study gave their personal impressions and perceptions about using VAM in teacher evaluation models. The main strategy for this analysis was surveys with a focus on individual teacher views about the use of value-added growth models as a part of the VAMIN teacher evaluation. This research occurred through surveys of multiple teachers teaching at VAMSD following a pre-scripted survey protocol of multiple questions regarding teacher perceptions towards the use of VAM in the VAMIN teacher evaluation model. Participants were selected through census sampling in that each eligible participant was prompted via an email letter (Appendix A) to fill out an online survey (Appendix B) on Qualtrics. Every effort was made to represent uniform diversity in gender, race, and educational attainment to gain multiple perspectives, but no specific criteria other than being a teacher in VAMSD during the 2013-14 school year was required. However, due to the nature of the research, time involvement, and willingness of potential participants to participate, it was difficult to represent all perspectives. Participation requests were sent via email to all MSDWT teachers through Microsoft Outlook.

Pilot Study

A pilot study of the survey was conducted by asking a focus group of five VAMSD administrators to take the survey. All focus group members were volunteers who received no compensation for their role in serving as a member of the group. Focus group members detailed strengths and weaknesses of the survey instrument and indicated which survey questions should be altered or eliminated. Additionally, they offered recommendations on how to improve the instrument.
Context

The surveys were completed at the participant’s discretion via computer at the location of their choice. Allowing participants anonymity and choice of location in completing surveys allowed participants to feel as comfortable as possible which likely yielded more meaningful and authentic responses. All survey responses were recorded via Qualtrics. The survey was open from April 15, 2014 through May 15, 2014. The desired goal was at least 100 survey responses. There were approximately 600 educators in the district.

Participant Roles

Each participant represented his or her beliefs about using VAM as part of a teacher-evaluation system through a survey of both closed and open ended questions. Each subject shared what hopes and fears the implementation of this program will had on them directly. They depicted the positives and negatives of evaluating teachers in this alternative way. Ultimately, the participant responses involved in this research project was the research. Survey responses of primarily Likert-based questions provided the raw data that was analyzed. Teacher responses provided an opportunity to see how teachers perceive this new method of evaluation in the state of Indiana.

Instrument

This research was conducted through a survey of closed and open ended questions. The participants answered a battery of survey questions that were designed to partially correspond to research done by CEEP (Cole et al., 2012). Permission to use their survey instrument was provided by Ms. Cole, the lead researcher via email correspondence. Surveys (Appendix B) were completed individually following a pre-established protocol to promote uniformity in the process for comparison purposes and theme identification. The survey protocol was structured in
a way to elicit details of teacher background, statistical expertise, knowledge of VAM, support of VAM in teacher evaluation models, perceived fairness in using VAM in teacher evaluation models, perceived student effects through using VAM, and perceived teacher effects through using VAM. Survey questions were both closed and open-ended with associated prompts to elicit meaningful and relevant teacher responses. All surveys were recorded and summarized on a computer using a computer program called Qualtrics.

**Analysis**

Survey questions were analyzed individually with the percentage of respondents who chose specific answers to each question detailed. After the survey window closed, the item responses were analyzed question by question. Each open-ended question with responses was categorized into a large generalized theme that captured the meaning of the responses. After all questions were categorized, an analysis of the overarching themes of the data was conducted to allow the responses to be put into the most salient themes. The responses were analyzed again, but in terms of variants of each theme. These themes and variants were used to make generalizations about participant beliefs on the use of VAM in teacher evaluation models.

These generalizations about participant responses occurred through triangulation of data. At least three questions and responses from the participants for each theme variant were documented to validate the use of specific variant themes to depict the group’s beliefs and perceptions.

**Validity**

The methods employed in this research follow the notion of researcher as detective, in that evidence was sought about cause and effects with careful consideration of causes with thorough examination of possible alternate causes. Additionally, low-inference descriptors were
used to describe participant responses with verbatim quotations as the primary form of open-ended questions. Triangulated data was used to understand and categorize responses. Finally, researcher reflexivity was utilized as a way to mitigate potential research-biases in drawing conclusions. The utilization of these methods made the results and conclusions plausible, credible, and trustworthy, thus making the research as valid as possible.

**Conclusion**

The intent of this research was to bring to light perceptions that teachers have about the use of value-added measures as part of evaluation models. This research gave voice to authentic teacher perceptions that could be used to improve value-added models for evaluation purposes. Additionally, this research has the potential to spur additional research which could shape future educational policy.
CHAPTER 4

RESULTS

The research conducted in this study is focused on investigating the impact of a value-added teacher evaluation model on teachers in a large urban Indianapolis school district. This research was conducted through five distinct investigations:

- **VETCS**: Is VAMIN resulting in better student achievement when compared to the previous year’s non-VAMIN teacher evaluation?
- **VGCS**: Is VAMIN resulting in better student growth when compared to the previous year’s non-VAMIN teacher evaluation?
- **VMRS**: Are there relationships between student achievement growth under a VAMIN-based teacher evaluation when controlling for student gender, race, special education status, socioeconomic status, Section 504 status, English learner status, and student previous year achievement score?
- **VCCS**: Is there consistency between teacher effectiveness ratings and principal-based observation evaluations?
- **VSTI**: What impacts do teachers perceive that VAMIN has on teachers in VAMSD?
Table 4.1 displays the mean scale score in ELA, math, science, and social studies for the 2011-2012 and 2012-2013 school years. Additionally, the respective grade level, number of students tested, mean scale score, mean difference between scale scores, degrees of freedom, T-value, and p-value are displayed. Finally, statistically significant mean differences are highlighted at an $\alpha = .05$ level.

In all grades except seventh in ELA, the mean score was higher in 2012-2013 than 2011-2012. Similarly, in all grades except sixth in math, the mean score was higher in 2012-2013 than 2011-2012. In social studies, all grades had a higher mean score in 2012-2013 than 2011-2012. This indicates that VAMIN may be playing a part in increasing student achievement in math. Conversely, the mean score in science was not higher in any grade levels in 2012-2013 compared to 2011-2012.

It appears that VAMIN may be impacting student ISTEP+ scores except science in a positive way. In order to determine if the score increases or decreases are statistically significant, t-tests were conducted. Prior to t-tests being conducted, Levene’s Test for Equal Variances was conducted to determine if an equal variance assumption should be used. The $\alpha$-level was set at .05.

Tables 4.1 also details the results of the t-tests. All tests were 2-tailed to determine if significant differences existed in either direction.
Table 4.1

### t-Test for Equality of Means for ELA, Math, Science, and Social Studies Scale Scores

<table>
<thead>
<tr>
<th>Subject</th>
<th>Grade</th>
<th>2012</th>
<th></th>
<th>2013</th>
<th></th>
<th>MD</th>
<th>df</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELA</td>
<td>3</td>
<td>811</td>
<td>461.67</td>
<td>803</td>
<td>464.88</td>
<td>3.207</td>
<td>1612</td>
<td>1.077</td>
<td>.282</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>744</td>
<td>473.91</td>
<td>771</td>
<td>482.90</td>
<td>8.989</td>
<td>1513</td>
<td>2.778</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>805</td>
<td>496.94</td>
<td>727</td>
<td>501.57</td>
<td>4.631</td>
<td>1528.474</td>
<td>1.682</td>
<td>.093</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>797</td>
<td>518.14</td>
<td>821</td>
<td>521.59</td>
<td>3.445</td>
<td>1616</td>
<td>.936</td>
<td>.349</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>785</td>
<td>537.04</td>
<td>823</td>
<td>534.27</td>
<td>-2.777</td>
<td>1600.348</td>
<td>-1.039</td>
<td>.299</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>737</td>
<td>542.76</td>
<td>774</td>
<td>551.86</td>
<td>9.097</td>
<td>9109</td>
<td>2.681</td>
<td>.007</td>
</tr>
<tr>
<td>Math</td>
<td>3</td>
<td>815</td>
<td>463.03</td>
<td>814</td>
<td>468.18</td>
<td>5.150</td>
<td>1596.733</td>
<td>1.339</td>
<td>.181</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>756</td>
<td>486.71</td>
<td>779</td>
<td>499.83</td>
<td>13.116</td>
<td>1533</td>
<td>3.586</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>813</td>
<td>530.60</td>
<td>738</td>
<td>537.92</td>
<td>7.323</td>
<td>1548.571</td>
<td>2.317</td>
<td>.021</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>803</td>
<td>536.59</td>
<td>830</td>
<td>533.11</td>
<td>-3.477</td>
<td>1631</td>
<td>-1.063</td>
<td>.288</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>785</td>
<td>548.20</td>
<td>833</td>
<td>565.26</td>
<td>17.053</td>
<td>1616</td>
<td>4.675</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>738</td>
<td>585.47</td>
<td>779</td>
<td>587.84</td>
<td>2.373</td>
<td>1514.051</td>
<td>.690</td>
<td>.490</td>
</tr>
<tr>
<td>Science</td>
<td>4</td>
<td>751</td>
<td>486.94</td>
<td>778</td>
<td>425.91</td>
<td>-61.028</td>
<td>1470.100</td>
<td>-22.700</td>
<td>.000</td>
</tr>
<tr>
<td>Social Studies</td>
<td>5</td>
<td>798</td>
<td>490.31</td>
<td>826</td>
<td>487.25</td>
<td>-3.063</td>
<td>1622</td>
<td>-.957</td>
<td>.339</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>809</td>
<td>503.13</td>
<td>734</td>
<td>509.57</td>
<td>6.435</td>
<td>1541</td>
<td>2.212</td>
<td>.027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>792</td>
<td>507.80</td>
<td>831</td>
<td>516.47</td>
<td>8.669</td>
<td>1591.289</td>
<td>3.177</td>
<td>.002</td>
</tr>
</tbody>
</table>

MD is Mean Difference
Shading indicates statistical significance at $\alpha = .05$

Table 4.1 indicates statistically significant increases in: ELA, grades 4 and 8; math, grades 4, 5, and 7; social studies, grades 5 and 7.

**Discussion**

VAMIN is not consistently resulting in better student achievement in VAMSD. Student mean test scores did not typically improve at a statistically significant level under VAMIN when compared to non-VAMIN test scores.

VAMIN resulted in statistically significant increases in seven out of 16 (43.75 percent) testing groups. This is an indication that the change in mean test scores was not due to chance in seven out of 16 instances. Hence, there is a reason for the changes in those seven areas.

However, those reasons have not been determined yet. Further investigation was conducted to
determine whether or not growth in student scale scores had similar results. The results of that investigation are listed in the next section.

**VGCS**

Statistics and Analysis

Table 4.2 displays the mean growth in scale scores from one grade level to the next in ELA and math in grades three through eight. Science and social studies were not included in this analysis as there were not consecutive grade-level tests in science and social studies. Additionally, the respective grade level span, number of students tested, mean scale score growth, mean difference between scale score growth, degrees of freedom, $T$-value, and $p$-value are listed. Finally, statistically significant mean differences are highlighted at an $\alpha = .05$ level.

In all grade spans except 5-6 for ELA, the mean growth in scores was higher in 2012-2013 than 2011-2012. This indicates that VAMIN may be playing a part in increasing student growth in ELA and math. Additionally, it appears that student growth is greater in math than ELA. Thus, VAMIN might have a greater impact on math growth than ELA growth. Again, this analysis is preliminary and causation cannot be implied at this point.

As indicated in the discussion above, it appears that VAMIN may be impacting growth in student ISTEP+ scores in a positive way. In order to determine if the growth score increases or decreases are statistically significant, $t$-tests were conducted. Prior to $t$-tests being conducted, Levene’s Test for Equal Variances was conducted to determine if an equal variance assumption should be used. The $\alpha$-level was set at .05.

Table 4.2 details the results of the $t$-tests. All tests were 2-tailed to determine if significant differences existed in either direction.
Table 4.2

1-Test for Equality of Means for Growth in ELA and Math Scale Scores

<table>
<thead>
<tr>
<th>Subject</th>
<th>Grade</th>
<th>2012</th>
<th>2013</th>
<th>MD</th>
<th>df</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELA</td>
<td>3-4</td>
<td>600</td>
<td>658</td>
<td>16.15</td>
<td>.974</td>
<td>1256</td>
<td>.437</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td>687</td>
<td>619</td>
<td>19.46</td>
<td>6.369</td>
<td>1304</td>
<td>3.447</td>
</tr>
<tr>
<td></td>
<td>5-6</td>
<td>665</td>
<td>684</td>
<td>24.28</td>
<td>21.70</td>
<td>1347</td>
<td>-1.103</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>658</td>
<td>698</td>
<td>10.00</td>
<td>14.98</td>
<td>1354</td>
<td>2.202</td>
</tr>
<tr>
<td></td>
<td>7-8</td>
<td>619</td>
<td>676</td>
<td>3.35</td>
<td>14.89</td>
<td>1293</td>
<td>5.758</td>
</tr>
<tr>
<td>Math</td>
<td>3-4</td>
<td>627</td>
<td>660</td>
<td>17.39</td>
<td>32.38</td>
<td>1285</td>
<td>14.982</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td>700</td>
<td>632</td>
<td>31.71</td>
<td>50.17</td>
<td>1330</td>
<td>18.460</td>
</tr>
<tr>
<td></td>
<td>5-6</td>
<td>678</td>
<td>690</td>
<td>-2.59</td>
<td>4.45</td>
<td>1319.875</td>
<td>3.047</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>660</td>
<td>699</td>
<td>15.55</td>
<td>30.41</td>
<td>1357</td>
<td>14.856</td>
</tr>
<tr>
<td></td>
<td>7-8</td>
<td>623</td>
<td>675</td>
<td>32.17</td>
<td>36.84</td>
<td>1296</td>
<td>4.668</td>
</tr>
</tbody>
</table>

MD is Mean Difference
Shading indicates statistical significance at $\alpha = .05$

Table 4.2 indicates statistically significant increases in ELA in three grade spans, 4-5, 6-7, and 7-8. In math, four grade spans, 3-4, 4-5, 6-7, and 7-8, showed statistically significant increases in growth.

**Discussion**

VAMIN is consistently resulting in better growth in student achievement. Student mean test scores did consistently improve at a statistically significant level under VAMIN when compared to non-VAMIN test scores.

After descriptive analysis, VAMIN resulted in statistically significant increases in seven out of 10 (70 percent) testing groups. This is an indication that VAMIN may play a part in improving student achievement growth. However, there could be many variables that could be impacting student growth as demonstrated in the analysis such as curricular or staffing changes.

Further investigation was conducted to determine the impact of known variables on the growth in student scale scores. The results of that investigation are listed in the next section.
Statistics and Analysis

Table 4.3 displays the correlation coefficient \((r)\), coefficient of variation \((R^2)\), adjusted \(R^2\), and standard error of the estimate for multiple regression equations with dependent variables of growth in scale scores from one grade level to the next in ELA and math. Science and social studies were not included in this analysis as there are not consecutive grade-level tests in science and social studies. Additionally, the respective grade level span is listed.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Grade</th>
<th>(r)</th>
<th>(R^2)</th>
<th>Adjusted (R^2)</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELA</td>
<td>3-4</td>
<td>.245</td>
<td>.060</td>
<td>.054</td>
<td>43.191</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td>.305</td>
<td>.093</td>
<td>.087</td>
<td>47.247</td>
</tr>
<tr>
<td></td>
<td>5-6</td>
<td>.235</td>
<td>.055</td>
<td>.050</td>
<td>54.103</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>.553</td>
<td>.306</td>
<td>.302</td>
<td>43.891</td>
</tr>
<tr>
<td></td>
<td>7-8</td>
<td>.295</td>
<td>.087</td>
<td>.081</td>
<td>49.233</td>
</tr>
<tr>
<td>Math</td>
<td>3-4</td>
<td>.440</td>
<td>.194</td>
<td>.189</td>
<td>45.667</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td>.421</td>
<td>.177</td>
<td>.172</td>
<td>50.077</td>
</tr>
<tr>
<td></td>
<td>5-6</td>
<td>.337</td>
<td>.113</td>
<td>.108</td>
<td>46.544</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>.249</td>
<td>.062</td>
<td>.057</td>
<td>60.585</td>
</tr>
<tr>
<td></td>
<td>7-8</td>
<td>.262</td>
<td>.069</td>
<td>.063</td>
<td>55.711</td>
</tr>
</tbody>
</table>

In all grade spans except 6-7 for ELA, the correlation coefficient \((r)\) is less than .5. This indicates that the variables in the models have weak to moderate correlation with growth in scale scores (Dancey, 2008). Additionally, the coefficients indicate that there are variables unaccounted for that have a greater impact on the variation in growth in student test scores than those included in the models.

An interesting occurrence is that \(R^2\) has a great deal of variability across grade spans and subjects. The values range from a low of .055 in sixth grade ELA to a high of .306 in seventh grade ELA. This indicates that other unaccounted for factors are playing a varied role in the
impact on student growth. One would expect that the variation would be more consistent across subjects and grade levels.

Tables 4.4 through 4.8 detail the results of the multiple regressions.

Table 4.4

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td></td>
</tr>
<tr>
<td>ELA (Constant)</td>
<td>90.503</td>
<td>16.847</td>
<td>5.372</td>
<td>.000</td>
</tr>
<tr>
<td>VAMIN</td>
<td>3.322</td>
<td>2.453</td>
<td>.037</td>
<td>1.355</td>
</tr>
<tr>
<td>Male</td>
<td>-4.347</td>
<td>2.496</td>
<td>-0.049</td>
<td>-1.742</td>
</tr>
<tr>
<td>White</td>
<td>11.207</td>
<td>3.329</td>
<td>.116</td>
<td>3.366</td>
</tr>
<tr>
<td>Non-SPED</td>
<td>11.179</td>
<td>3.994</td>
<td>.079</td>
<td>2.799</td>
</tr>
<tr>
<td>Non-FRL</td>
<td>11.427</td>
<td>3.203</td>
<td>.126</td>
<td>3.568</td>
</tr>
<tr>
<td>Non-504</td>
<td>4.192</td>
<td>12.606</td>
<td>.009</td>
<td>.333</td>
</tr>
<tr>
<td>Non-EL</td>
<td>6.264</td>
<td>3.631</td>
<td>.054</td>
<td>1.725</td>
</tr>
<tr>
<td>Prior ELA</td>
<td>-2.18</td>
<td>.027</td>
<td>-.278</td>
<td>-8.011</td>
</tr>
<tr>
<td>R² = .060</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math (Constant)</td>
<td>140.017</td>
<td>15.955</td>
<td>8.775</td>
<td>.000</td>
</tr>
<tr>
<td>VAMIN</td>
<td>15.706</td>
<td>2.564</td>
<td>.155</td>
<td>6.126</td>
</tr>
<tr>
<td>Male</td>
<td>-4.158</td>
<td>2.604</td>
<td>-.041</td>
<td>-1.597</td>
</tr>
<tr>
<td>White</td>
<td>14.755</td>
<td>3.471</td>
<td>.134</td>
<td>4.251</td>
</tr>
<tr>
<td>Non-SPED</td>
<td>13.240</td>
<td>4.142</td>
<td>.082</td>
<td>3.197</td>
</tr>
<tr>
<td>Non-FRL</td>
<td>9.539</td>
<td>3.354</td>
<td>.092</td>
<td>2.844</td>
</tr>
<tr>
<td>Non-504</td>
<td>4.065</td>
<td>13.331</td>
<td>.008</td>
<td>.305</td>
</tr>
<tr>
<td>Non-EL</td>
<td>-2.092</td>
<td>3.598</td>
<td>-.016</td>
<td>-.581</td>
</tr>
<tr>
<td>Prior ELA</td>
<td>-.305</td>
<td>.020</td>
<td>-.457</td>
<td>-15.340</td>
</tr>
<tr>
<td>R² = .194</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SE is Standard Error
Shading indicates statistical significance at α = .05
Table 4.5

Multiple Regression Model Summary for 5th Grade Growth in ELA and Math Scores

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td>t</td>
</tr>
<tr>
<td>ELA</td>
<td>(Constant)</td>
<td>117.115</td>
<td>16.881</td>
<td>6.938</td>
</tr>
<tr>
<td></td>
<td>VAMIN</td>
<td>12.542</td>
<td>2.641</td>
<td>.127</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>-8.997</td>
<td>2.671</td>
<td>-.091</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>8.874</td>
<td>3.533</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td>Non-SPED</td>
<td>16.722</td>
<td>4.284</td>
<td>.106</td>
</tr>
<tr>
<td></td>
<td>Non-FRL</td>
<td>8.653</td>
<td>3.341</td>
<td>.086</td>
</tr>
<tr>
<td></td>
<td>Non-504</td>
<td>4.948</td>
<td>10.938</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>Non-EL</td>
<td>2.556</td>
<td>4.125</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>Prior ELA</td>
<td>-.265</td>
<td>.028</td>
<td>-.304</td>
</tr>
<tr>
<td></td>
<td>$R^2 = .093$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>(Constant)</td>
<td>149.994</td>
<td>16.586</td>
<td>9.043</td>
</tr>
<tr>
<td></td>
<td>VAMIN</td>
<td>20.015</td>
<td>2.779</td>
<td>.182</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>5.974</td>
<td>2.797</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>4.633</td>
<td>3.726</td>
<td>.039</td>
</tr>
<tr>
<td></td>
<td>Non-SPED</td>
<td>28.656</td>
<td>4.418</td>
<td>.166</td>
</tr>
<tr>
<td></td>
<td>Non-FRL</td>
<td>15.526</td>
<td>3.546</td>
<td>.138</td>
</tr>
<tr>
<td></td>
<td>Non-504</td>
<td>-4.674</td>
<td>11.591</td>
<td>-.010</td>
</tr>
<tr>
<td></td>
<td>Non-EL</td>
<td>1.731</td>
<td>4.100</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>Prior ELA</td>
<td>-.311</td>
<td>.024</td>
<td>-.389</td>
</tr>
<tr>
<td></td>
<td>$R^2 = .177$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SE is Standard Error

Shading indicates statistical significance at $\alpha = .05$
Table 4.6

*Multiple Regression Model Summary for 6th Grade Growth in ELA and Math Scores*

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td></td>
</tr>
<tr>
<td>ELA (Constant)</td>
<td>-31.393</td>
<td>27.076</td>
<td>-1.159</td>
<td>.246</td>
</tr>
<tr>
<td>VAMIN</td>
<td>3.706</td>
<td>2.959</td>
<td>.033</td>
<td>1.253</td>
</tr>
<tr>
<td>Male</td>
<td>-4.841</td>
<td>2.980</td>
<td>-.044</td>
<td>-1.624</td>
</tr>
<tr>
<td>White</td>
<td>10.894</td>
<td>3.832</td>
<td>.092</td>
<td>2.843</td>
</tr>
<tr>
<td>Non-SPED</td>
<td>27.712</td>
<td>5.118</td>
<td>.146</td>
<td>5.414</td>
</tr>
<tr>
<td>Non-FRL</td>
<td>12.598</td>
<td>3.662</td>
<td>.112</td>
<td>3.440</td>
</tr>
<tr>
<td>Non-504</td>
<td>16.248</td>
<td>22.251</td>
<td>.019</td>
<td>.730</td>
</tr>
<tr>
<td>Non-EL</td>
<td>-.609</td>
<td>4.993</td>
<td>-.004</td>
<td>-.122</td>
</tr>
<tr>
<td>Prior ELA</td>
<td>.005</td>
<td>.034</td>
<td>.005</td>
<td>.146</td>
</tr>
<tr>
<td>R² = .055</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Math (Constant) | 125.587 | 22.308 | 5.630 | .000 |
| VAMIN  | 6.781 | 2.520 | .069 | 2.690 | .007 |
| Male   | .272 | 2.552 | .003 | .107 | .915 |
| White  | 18.935 | 3.240 | .180 | 5.843 | .000 |
| Non-SPED | 16.038 | 4.311 | .097 | 3.721 | .000 |
| Non-FRL | 15.814 | 3.163 | .159 | 5.000 | .000 |
| Non-504 | -17.259 | 19.144 | -.023 | -.902 | .367 |
| Non-EL  | .028 | 4.021 | .000 | .007 | .994 |
| Prior ELA | -.264 | .023 | -.359 | -11.553 | .000 |
| R² = .113 |       |       |       |       |       |

SE is Standard Error
Shading indicates statistical significance at α = .05
Table 4.7

Multiple Regression Model Summary for 7th Grade Growth in ELA and Math Scores

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td>(Constant)</td>
<td>228.014</td>
<td>21.955</td>
<td>10.385</td>
</tr>
<tr>
<td></td>
<td>VAMIN</td>
<td>6.410</td>
<td>2.391</td>
<td>.061</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>-3.241</td>
<td>2.422</td>
<td>-0.031</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>7.073</td>
<td>3.004</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>Non-SPED</td>
<td>17.766</td>
<td>4.462</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>Non-FRL</td>
<td>12.842</td>
<td>2.847</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>Non-504</td>
<td>-9.405</td>
<td>19.681</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>Non-EL</td>
<td>-3.795</td>
<td>4.300</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>Prior ELA</td>
<td>-0.439</td>
<td>0.200</td>
<td>-0.608</td>
</tr>
<tr>
<td></td>
<td>R² = .306</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>(Constant)</td>
<td>10.147</td>
<td>31.068</td>
<td>0.327</td>
</tr>
<tr>
<td></td>
<td>VAMIN</td>
<td>24.385</td>
<td>3.293</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>-4.079</td>
<td>3.309</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>8.525</td>
<td>4.147</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>Non-SPED</td>
<td>18.932</td>
<td>5.963</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>Non-FRL</td>
<td>13.081</td>
<td>3.956</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>Non-504</td>
<td>22.571</td>
<td>27.163</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>Non-EL</td>
<td>-9.879</td>
<td>5.740</td>
<td>-0.048</td>
</tr>
<tr>
<td></td>
<td>Prior ELA</td>
<td>-0.077</td>
<td>0.030</td>
<td>-0.080</td>
</tr>
<tr>
<td></td>
<td>R² = .062</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SE is Standard Error

Shading indicates statistical significance at α = .05
Table 4.8

*Multiple Regression Model Summary for 8th Grade Growth in ELA and Math Scores*

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>β</td>
<td>t</td>
</tr>
<tr>
<td>ELA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>-70.906</td>
<td>23.775</td>
<td>-2.982</td>
<td>.003</td>
</tr>
<tr>
<td>VAMIN</td>
<td>16.202</td>
<td>2.749</td>
<td>.158</td>
<td>5.895</td>
</tr>
<tr>
<td>Male</td>
<td>-8.133</td>
<td>2.784</td>
<td>-.079</td>
<td>-2.921</td>
</tr>
<tr>
<td>White</td>
<td>17.505</td>
<td>3.469</td>
<td>.164</td>
<td>5.046</td>
</tr>
<tr>
<td>Non-SPED</td>
<td>8.567</td>
<td>5.092</td>
<td>.047</td>
<td>1.682</td>
</tr>
<tr>
<td>Non-FRL</td>
<td>7.952</td>
<td>3.376</td>
<td>.077</td>
<td>2.356</td>
</tr>
<tr>
<td>Non-504</td>
<td>61.002</td>
<td>16.544</td>
<td>.099</td>
<td>3.687</td>
</tr>
<tr>
<td>Non-EL</td>
<td>2.587</td>
<td>5.751</td>
<td>.013</td>
<td>1.450</td>
</tr>
<tr>
<td>Prior ELA</td>
<td></td>
<td></td>
<td>-.013</td>
<td>-.384</td>
</tr>
<tr>
<td>R² = .087</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>50.672</td>
<td>23.343</td>
<td>2.171</td>
<td>.030</td>
</tr>
<tr>
<td>VAMIN</td>
<td>10.893</td>
<td>3.103</td>
<td>.095</td>
<td>3.510</td>
</tr>
<tr>
<td>Male</td>
<td>-6.007</td>
<td>3.124</td>
<td>-.052</td>
<td>-1.923</td>
</tr>
<tr>
<td>White</td>
<td>8.676</td>
<td>3.998</td>
<td>.073</td>
<td>2.170</td>
</tr>
<tr>
<td>Non-SPED</td>
<td>15.665</td>
<td>5.733</td>
<td>.076</td>
<td>2.733</td>
</tr>
<tr>
<td>Non-FRL</td>
<td>2.675</td>
<td>3.822</td>
<td>.023</td>
<td>.700</td>
</tr>
<tr>
<td>Non-504</td>
<td>68.907</td>
<td>18.721</td>
<td>.099</td>
<td>3.681</td>
</tr>
<tr>
<td>Non-EL</td>
<td>1.575</td>
<td>6.147</td>
<td>.007</td>
<td>.256</td>
</tr>
<tr>
<td>Prior ELA</td>
<td>-.199</td>
<td>.027</td>
<td>-.244</td>
<td>-7.281</td>
</tr>
<tr>
<td>R² = .069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SE is Standard Error

Shading indicates statistical significance at $\alpha = .05$

**Discussion**

Overall, student achievement growth is most impacted by student previous year achievement score with a negative association between the two variables. Additionally, VAMIN does not have a consistently strong impact on student achievement growth.

After thorough analysis, two important results are apparent. The first is that in all models except for one (90 percent), VAMIN does not have the largest standardized coefficient in magnitude. This means that other tested variables have a greater impact on student test scores than using a value-added teacher evaluation.
The second important result is that in all models except for three (70 percent), prior-year test score has the largest standardized coefficient in magnitude. This means that out of all the variables analyzed in the multiple regression equations, prior test score routinely had the greatest impact on student growth. Additionally, the coefficient is negative in all models except for grade six ELA, and even there the coefficient was essentially zero with a value of .005. Since the coefficient was negative, this means that for each one-point increase in a student’s prior year score, that student’s growth would decrease at the coefficient value. This indicates that students who started with low prior-year scores were likely to show more growth than students with a high prior-year score.

Now that it has been shown that VAMIN does not have a strong impact on student growth and prior-year test score does, further analysis was conducted to determine whether or not VAMIN and traditional teacher observation measures correlate. The results of that investigation are listed in the next section.

VCCS

Statistics and Analysis

In a correlation analysis between teacher VAMIN Score and Teacher Effectiveness Ratings (TER) score of \( n = 149 \) teachers, the correlation coefficient was calculated at \( r = .196 \) with a 2-tailed significance level of \( p = .017 \).

Discussion

There is very poor consistency between TER which are a form of student growth percentiles and principal-based observation evaluations. The results indicate that the correlation is significant, but it is a very weak positive association. This means that evaluations under VAMIN and TER scores are weakly related, or the scores do not equate linearly very well. So
these scores do not exhibit related movements consistently. As one increases the other does not necessarily increase at regular intervals.

Since it was shown that VAMIN and TER are very weakly correlated, further analysis was conducted to determine teacher perceptions of VAMIN. The results of that investigation are listed in the next section.

**VSTI**

**Statistics and Analysis**

Tables A.1 through A.8 (Appendix D) display participant demographic and background information responses. Tables 4.9 through 4.15 detail participant responses on participant perceptions of value-added teacher evaluations.

Out of 532 possible participants, 97 teachers agreed to complete the survey. Several possible participants (38) opened the survey and decided to discontinue participation. Additionally, many participants did not complete all questions in the survey. Potentially, this is due to a feeling of distrust surrounding value-added teacher evaluations.

Generally, the survey participants were representative of VAMSD teachers (teachers who taught in VAMSD during the 2013-2014 school year). The typical participant was a white female between the ages of 21 and 40. She had a Master’s Degree with 10 plus years of experience. She taught at the high school level in a subject other than math or English, and had never held a full-time position in a field other than education.

Table 4.9 indicates that the statement with the highest mean value was that teacher effectiveness impacts student achievement at 4.13. Additionally, two other statements also had high mean values. An effective teacher evaluation system informs professional development had
a mean of 3.87, and an effective teacher evaluation system drives professional development had mean 3.60.

The statement VAM-based teacher evaluations have resulted in improved teaching and learning had the lowest mean at 2.24. Additionally, the statement teacher evaluation should be tied to compensation also had a low mean of 2.47.

Table 4.9 also indicated an interesting pattern of responses in that nearly half of the respondents (48 percent) disagreed with the statement I believe that teacher evaluation should be linked to student growth.

These responses show that teachers believe that teachers are important in student learning, but they do not believe that VAM-based teacher evaluations are beneficial aside from professional development. Also, they do not believe that teacher evaluations should be tied to student growth.
Table 4.9

Survey Question: For each statement below, please indicate your level of agreement.

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>SD</th>
<th>D</th>
<th>NAD</th>
<th>A</th>
<th>SA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I believe that teacher effectiveness affects student achievement.</td>
<td>1 (1%)</td>
<td>3 (4%)</td>
<td>5 (7%)</td>
<td>36 (53%)</td>
<td>23 (34%)</td>
<td>4.13</td>
</tr>
<tr>
<td>2</td>
<td>I believe that student achievement can be validly measured.</td>
<td>4 (6%)</td>
<td>14 (21%)</td>
<td>15 (22%)</td>
<td>32 (47%)</td>
<td>3 (4%)</td>
<td>3.24</td>
</tr>
<tr>
<td>3</td>
<td>I believe that student academic growth can be validly measured.</td>
<td>3 (4%)</td>
<td>12 (18%)</td>
<td>11 (16%)</td>
<td>34 (50%)</td>
<td>8 (12%)</td>
<td>3.47</td>
</tr>
<tr>
<td>4</td>
<td>I believe that teacher evaluation should be linked to student growth.</td>
<td>14 (21%)</td>
<td>18 (27%)</td>
<td>12 (18%)</td>
<td>21 (31%)</td>
<td>2 (3%)</td>
<td>2.69</td>
</tr>
<tr>
<td>5</td>
<td>I believe that instruction can be accurately evaluated and judged.</td>
<td>6 (9%)</td>
<td>13 (19%)</td>
<td>11 (16%)</td>
<td>34 (50%)</td>
<td>4 (6%)</td>
<td>3.25</td>
</tr>
<tr>
<td>6</td>
<td>I believe that the relationship between teaching and learning can be accurately applied to an evaluation of teaching.</td>
<td>5 (7%)</td>
<td>20 (30%)</td>
<td>15 (22%)</td>
<td>24 (36%)</td>
<td>3 (4%)</td>
<td>3.00</td>
</tr>
<tr>
<td>7</td>
<td>Prior to the use of VAM, the teacher evaluation processes in Indiana needed improvement.</td>
<td>4 (6%)</td>
<td>13 (19%)</td>
<td>18 (27%)</td>
<td>24 (36%)</td>
<td>8 (12%)</td>
<td>3.28</td>
</tr>
<tr>
<td>8</td>
<td>An effective teacher evaluation system informs professional development.</td>
<td>2 (3%)</td>
<td>8 (12%)</td>
<td>4 (6%)</td>
<td>37 (54%)</td>
<td>17 (25%)</td>
<td>3.87</td>
</tr>
<tr>
<td>9</td>
<td>An effective teacher evaluation system drives professional development.</td>
<td>2 (3%)</td>
<td>13 (19%)</td>
<td>7 (10%)</td>
<td>34 (50%)</td>
<td>12 (18%)</td>
<td>3.60</td>
</tr>
<tr>
<td>10</td>
<td>VAM-based teacher evaluations have resulted in improved teaching and learning.</td>
<td>22 (32%)</td>
<td>18 (26%)</td>
<td>18 (26%)</td>
<td>10 (15%)</td>
<td>0 (0%)</td>
<td>2.24</td>
</tr>
<tr>
<td>11</td>
<td>Teacher evaluation should be tied to compensation.</td>
<td>20 (29%)</td>
<td>15 (22%)</td>
<td>16 (24%)</td>
<td>15 (22%)</td>
<td>2 (3%)</td>
<td>2.47</td>
</tr>
</tbody>
</table>

SD is Strongly Disagree, D is Disagree, NAD is Neither Agree nor Disagree, A is Agree, SA is Strongly Agree, MA is Mean Agreement

Table 4.10 indicates that VAMSD teachers believe that they are somewhat familiar with Public Law 90 which governs teacher evaluation requirements. While the questions did not ask respondents about their perception of the merits of the law a few expressed great displeasure by stating:

I am aware of the PL 90, and strongly believe it is necessary, but am highly disappointed that so many crucial variables are completely disregarded. i.e.: socio-economic factors of the school clientele, parental involvement, student attendance and discipline. The
evaluation, as it is now, is not an effective technique to strengthen a teacher's needs, and threatens the jobs of the very people who are trying to raise the scores of their students by whatever means available to them. \textit{(sic)}

Feelings of anger and resentment were again expressed in the following statement:

When our political leaders quietly slipped through legislation that was not represented in their campaign and was not sufficiently covered by our media, and was not debated or review, their lack of transparency and the subsequent backlash among teachers ALL OVER THE STATE has revealed that their priority is $ NOT $ the improvement of public education, and their fabricated "education crisis" makes for a good soundbyte but has only eroded the quality work many teachers were voluntarily supplying to their communities--hence this survey. \textit{amiright? (sic)}

Table 4.10

Survey Question: How familiar are you with Public Law 90 that required the implementation of a new teacher evaluation system starting in the 2012-13 school year?

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Extremely familiar</td>
<td>5</td>
<td>7%</td>
</tr>
<tr>
<td>4</td>
<td>Adequately familiar</td>
<td>25</td>
<td>37%</td>
</tr>
<tr>
<td>3</td>
<td>Somewhat familiar</td>
<td>21</td>
<td>31%</td>
</tr>
<tr>
<td>2</td>
<td>Slightly familiar</td>
<td>9</td>
<td>13%</td>
</tr>
<tr>
<td>1</td>
<td>Not at all familiar</td>
<td>8</td>
<td>12%</td>
</tr>
</tbody>
</table>

Mean Agreement = 2.85

Table 4.11 indicates that VAMSD teachers have largely learned about the requirements of Public Law 90 through discussions with administrators or other teachers. Similar to the prior questions, participants felt so frustrated by Public Law 90 that they stated:

I skip breakfast and lunch daily. I come in on weekends and stay late evenings WORKING NON-STOP. The added trainings, meetings, and requirements have all taken away from the multiple positive things I was doing as a PROFESSIONAL because I
CARE about my students and classes. I barely have time to do the work of the teacher b/c I'm hoop-jumping and documenting all of my hoop-jumping. NO OTHER PROFESSION DOES THIS. It is not sustainable and the consequences are DIRE. (sic)

Table 4.11

Survey Question: How have you become familiar with the requirements of Public Law 90? (Select all that apply)

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Read the legislation</td>
<td>16</td>
<td>26%</td>
</tr>
<tr>
<td>2</td>
<td>Attended workshops/seminars</td>
<td>17</td>
<td>27%</td>
</tr>
<tr>
<td>3</td>
<td>Participated in webinars</td>
<td>4</td>
<td>6%</td>
</tr>
<tr>
<td>4</td>
<td>Had discussions with administrators</td>
<td>41</td>
<td>66%</td>
</tr>
<tr>
<td>5</td>
<td>Discussed requirements of the law with other teachers</td>
<td>49</td>
<td>79%</td>
</tr>
<tr>
<td>6</td>
<td>Spoke with IDOE officials or reviewed information on the IDOE website</td>
<td>15</td>
<td>24%</td>
</tr>
</tbody>
</table>

Total Respondents = 62

Table 4.12 indicates that teachers believe that teachers, teacher association leadership, principals, and central office staff have been the most involved in developing VAM-based teacher evaluations in VAMSD.

Table 4.12

Survey Question: Which stakeholder groups that you know of have been or will be a part of your district’s VAM-based teacher evaluation development process? (Select all that apply)

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Parents</td>
<td>12</td>
<td>18%</td>
</tr>
<tr>
<td>2</td>
<td>Students</td>
<td>8</td>
<td>12%</td>
</tr>
<tr>
<td>3</td>
<td>Teachers</td>
<td>43</td>
<td>65%</td>
</tr>
<tr>
<td>4</td>
<td>Teacher association leadership</td>
<td>38</td>
<td>58%</td>
</tr>
<tr>
<td>5</td>
<td>Principals</td>
<td>53</td>
<td>80%</td>
</tr>
<tr>
<td>6</td>
<td>Central office staff</td>
<td>45</td>
<td>68%</td>
</tr>
<tr>
<td>7</td>
<td>Technology personnel</td>
<td>8</td>
<td>12%</td>
</tr>
<tr>
<td>8</td>
<td>Community members</td>
<td>4</td>
<td>6%</td>
</tr>
<tr>
<td>9</td>
<td>Others</td>
<td>6</td>
<td>9%</td>
</tr>
</tbody>
</table>

Total Respondents = 66

Table 4.13 indicates that six questions have a high mean concern for VAMSD teachers: resources to provide training for staff, resources for increased compensation, building the
capacity for understanding among school personnel, communication to key stakeholders, ongoing support for professional development, and clear guidance concerning the interpretation of new law.

Table 4.13

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>NC</th>
<th>SC</th>
<th>SWC</th>
<th>VC</th>
<th>EC</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Resources to conduct classroom observations</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>12</td>
<td>9</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(23%)</td>
<td>(21%)</td>
<td>(23%)</td>
<td>(19%)</td>
<td>(15%)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Resources to collect student performance data</td>
<td>8</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(13%)</td>
<td>(23%)</td>
<td>(23%)</td>
<td>(21%)</td>
<td>(20%)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Resources to provide training for evaluators</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>12</td>
<td>20</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(10%)</td>
<td>(16%)</td>
<td>(23%)</td>
<td>(19%)</td>
<td>(32%)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Resources to provide training for staff</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>14</td>
<td>21</td>
<td>3.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5%)</td>
<td>(13%)</td>
<td>(23%)</td>
<td>(23%)</td>
<td>(35%)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Resources for increased compensation</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>16</td>
<td>26</td>
<td>3.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6%)</td>
<td>(10%)</td>
<td>(16%)</td>
<td>(26%)</td>
<td>(42%)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Building the capacity for understanding among school personnel</td>
<td>4</td>
<td>4</td>
<td>19</td>
<td>21</td>
<td>14</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6%)</td>
<td>(6%)</td>
<td>(31%)</td>
<td>(34%)</td>
<td>(23%)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Communication to key stakeholders</td>
<td>4</td>
<td>8</td>
<td>15</td>
<td>15</td>
<td>20</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6%)</td>
<td>(13%)</td>
<td>(24%)</td>
<td>(24%)</td>
<td>(32%)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>On-going support for professional development</td>
<td>2</td>
<td>5</td>
<td>17</td>
<td>15</td>
<td>23</td>
<td>3.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3%)</td>
<td>(8%)</td>
<td>(27%)</td>
<td>(24%)</td>
<td>(37%)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Clear guidance concerning the interpretation of new law</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>18</td>
<td>27</td>
<td>3.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6%)</td>
<td>(6%)</td>
<td>(15%)</td>
<td>(29%)</td>
<td>(44%)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Alignment of new law with policy</td>
<td>4</td>
<td>7</td>
<td>21</td>
<td>15</td>
<td>15</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6%)</td>
<td>(11%)</td>
<td>(34%)</td>
<td>(24%)</td>
<td>(24%)</td>
<td></td>
</tr>
</tbody>
</table>

NC is Not at all concerned, SC is Slightly concerned, SWC is Somewhat concerned, VC is Very concerned, EC is extremely concerned, MC is Mean concern

Table 4.14 indicates that many teachers have considered leaving the teaching profession. Of the 47 participants that said they have considered leaving the teaching profession, 47% of those teachers have taught for 10 or more years and 75% have taught for five or more years.

This suggests that many teachers are experiencing burn-out as teachers with less than five years of experience were less likely to say they have considered leaving the teaching profession.

Participants expressed a common theme in their reasons for considering leaving the teaching profession expressed by stating:
The manner in which I work requires an extraordinary level of energy (emotional and physical). I am still fairly young...but I do question whether I can keep up this pace for 20 more years. (sic)

Further concern was expressed with the following statement:

My career satisfaction was extremely high until 2 years ago. I am holding on in hopes that the powers that be will rescind this ineffective practice and realize that one size does not fit all and that students and teachers are indeed more than "just a test score." (sic)

Another participant responded in the following way:

When the veterans you love and admire who provide guidance, stability, and encouragement for the "long-haul" begin leaving MID-SEMESTER-- When the newbies stop smiling and chirping under the impossible burden of processing all the information and requirements--when the rock-solid professionals who NEVER COMPLAIN begin to show cracks in their foundation as the stress, disgust, and mistreatment extinguishes their light and oppresses their souls, you consider leaving the profession you LOVE, the profession to which you've devoted your LIFE, HEART, SOUL. YES. But you already knew that; that's why you posed the question. (sic)

All of these responses show a deep emotional hurt and betrayal tied to the implementation of value-added teacher evaluations that respondents claim was not present before the implementation of the new evaluation measures. These responses demonstrate teacher dissatisfaction with their work and anecdotal evidence that teachers are leaving the teaching profession due to policies mandating the use of value-added teacher evaluations.

Participant responses to the question have you considered leaving the teaching profession specifically because of the new evaluation system confirms the participant responses to the
previous question that teachers have largely considered leaving the teaching profession due to the use of a value-added teacher evaluation. Of the 36 participants that said they have considered leaving the teaching profession specifically because of the new evaluation system, 53% of those teachers have taught for 10 or more years and 77% have taught for five or more years. Participants expressed their dissatisfaction of VAM-based evaluations by stating, “I am judged based on students with very high absence rates, including truancy.” and

Only because the type of raise earned for being highly effective and/or effective is either distributed as a stipend and not attached to my base pay. Another reason is that the raises are lower than the cost of living each year. Find it very hard to increase salary with current evaluation system even if one earns effective or highly effective evaluations. (sic)

Also indicated in Table 4.14, is the response that many teachers would not recommend the teaching profession to prospective employees. Participants again expressed their dissatisfaction with what the teaching profession has become by stating, “NO. I have discouraged [sic] own children from entering the profession,” and “The business model to teaching model has not improved teaching. It has only haphazardly forced educators to narrow their choices of material to teach to students.”

Finally, responses indicated that teacher’s willingness to help fellow teachers has decreased due to the new teacher evaluation, but many teachers still expressed a willingness to help fellow teachers. Teachers illustrated these feelings through responses like, “Fellow teachers are in the trenches together and they all deserve any help I may be able to offer,” and “Opposite. I feel a greater responsibility to work with my team to maximize academic achievement,” and “Screw [sic] your new evaluations. I collaborate with my peers because it is the way things work.”
Table 4.14

Survey Question Responses Regarding Teaching Profession Perspectives

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you ever considered leaving the teaching profession?</td>
<td>47</td>
<td>16</td>
</tr>
<tr>
<td>(75%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have you considered leaving the teaching profession specifically because of the new evaluation system?</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td>(57%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Would you recommend the teaching profession to prospective employees?</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td>(37%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has your willingness to help fellow teachers decreased due to the new teacher evaluation?</td>
<td>21</td>
<td>42</td>
</tr>
<tr>
<td>(33%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.15 indicates that teachers overwhelmingly feel pressured to raise student test scores as the new evaluation rates teacher effectiveness on growth in student test scores.

Participants conveyed their feelings through responses like, “Not raising scores can directly effect [sic] my income, or the fact that I have a job,” and “I am told to teach to the test. Totally against my principles and what I was educated to do as a teacher,” (sic) and

Students openly state, ‘I can ruin my teacher's salary if I do poorly on this test’ followed by laughter. Colleagues report with fear and sorrow that they are not able to hold students accountable in their classroom because the student has been given all the power, and they know it. False accusations, time-wasters, and complaining have taken over the landscape of a system trying to individualize to remain "competitive" with schools who don't have to meet any of the requirements place on the public education system.” (sic)

Teachers also responded to a large extent that they have not witnessed cheating on high-stakes testing even though they felt pressured to raise student test scores.

These teachers are split as to whether or not student/teacher cheating is of greater concern due to the new teacher evaluation. However, teacher’s did express concern about cheating due to
the new teacher evaluation by stating, “The whole system is set up to encourage cheating” and “But you never know what someone might do under the pressure.”

<table>
<thead>
<tr>
<th>Table 4.15</th>
</tr>
</thead>
</table>

Survey Question Responses Regarding Perspectives on Cheating

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you feel pressured to raise student test scores?</td>
<td>54 (86%)</td>
<td>9 (14%)</td>
</tr>
<tr>
<td>Have you ever witnessed cheating on high stakes testing?</td>
<td>18 (29%)</td>
<td>45 (71%)</td>
</tr>
<tr>
<td>Is student/teacher cheating of greater concern to you due to the new teacher evaluation?</td>
<td>31 (50%)</td>
<td>31 (50%)</td>
</tr>
</tbody>
</table>

Discussion

Teachers generally perceive impacts from VAM-based teacher evaluations as negative. They report high levels of pressure to raise student test scores while simultaneously being demoralized to the point of considering leaving the teaching profession.

Also, unlike the 2007 Phi Delta Kappa/Gallup Poll and 2012 CEEP survey, teacher survey participants did not demonstrate agreement (34.3 percent) that student growth measures should be used for teacher evaluations. This is in stark contrast to the 82 percent of public participant supporters of growth measure use in teacher evaluations and 88.8 percent of superintendents.

Additionally, teacher participants indicated strong agreement with Nichols and Berliner (2007) that student/teacher cheating has become increasingly concerning through the adoption of VAM-based teacher evaluations. This is likely because teachers felt tremendous pressure to raise student test scores with 86 percent of respondents indicating that they do feel pressured to raise student test scores.
These results correspond with prior research (Anderman, Anderman, Yough, & Gimbert, 2010; Baker et al, 2010; Darling-Hammond, Wise, & Pease, 1983; MetLife, 2013 National Council for Teacher Quality, 2011; Whiteman, Shi, & Plucker, 2011) that VAM-based teacher evaluations result in teacher demoralization in that 57 percent of participants indicated that they have considered leaving the teaching profession due to the new evaluation system. Additionally, 63 percent of respondents would not recommend the teaching profession to prospective teachers.

Another negative consequence of VAM-based teacher evaluations is a reduction in teacher collaboration (Baker, et al., 2010). This was confirmed in the survey with 33 percent of survey respondents indicating that they were less willing to help fellow teachers due to the new evaluation.

Summary of Results

The following is a brief summary of results identified through the five research analyses of VAMSD’s usage of VAMIN detailed in this chapter.

**VETCS**

VAMIN is not consistently resulting in better student achievement in VAMSD. Student mean test scores did not regularly improve at a statistically significant level under VAMIN when compared to non-VAMIN test scores.

**VGCS**

VAMIN is consistently resulting in better growth in student achievement. Student mean test scores did frequently improve at a statistically significant level under VAMIN when compared to non-VAMIN test scores.
VMRS

Student achievement growth is most impacted by student previous year achievement score with a negative association. Additionally, VAMIN does not have a consistently strong impact on student achievement growth.

VCCS

There is very poor consistency between TER which are a form of student growth percentiles and principal-based observation evaluations. The results indicate that the correlation is significant, but it is a very weak positive association.

VSTI

Teachers generally perceive impacts from VAM-based teacher evaluations as negative. They report high levels of pressure to raise student test scores while simultaneously being demoralized to the point of considering leaving the teaching profession. VAMSD teachers do not believe that student growth measures should be used for teacher evaluations. VAMSD teachers believe that student/teacher cheating has become increasingly concerning through the adoption of VAM-based teacher evaluations, and they believe that VAM-based teacher evaluations have resulted in a reduction in teacher collaboration.

Implications of these findings are discussed in chapter 5.
CHAPTER 5

CONCLUSIONS

Summary

Context

In the fall of 2012, Indiana schools began transitioning to a new teacher evaluation model that evaluates teachers partially on value-added measures (VAM). These measures are a collection of statistical formulas and techniques that try to derive teacher effects on student performance. Indiana’s new evaluation model at the time of this writing requires the use of growth measures to evaluate teacher effectiveness which is a form of VAM. Any teacher that negatively affects student achievement or growth as indicated by VAM, through the interpretation and guidance of the Indiana Department of Education, cannot receive a pay increase and could be terminated depending on teaching experience and past evaluation ratings (Indiana Department of Education, 2012). VAM is related to test based-accountability which is a result of the No Child Left Behind Act of 2001. It mandated standardized testing to report on student progress through the use of school-based value-added models for accountability purposes (Olson, 2004). The switch to teacher-based value-added models was a result of the 2012 federal Race to the Top grant program which required the use of teacher affected student growth measures to document the number of effective and highly effective teachers in a school (U.S. Department of Education, 2012). Additionally, the push for using VAM is that it is believed that
value-added modeling can, “capture how much students learn during the school year, thereby
putting teachers on a more level playing field as they aim for tenure or additional pay” (David,
2010, p. 81). However, there is already some concern from educators and researchers that these
claims may not be true, especially considering Indiana’s interpretation of VAM in which student
growth percentiles are used to calculate teacher effectiveness in only grades three through eight
(Cavanaugh, 2011; Cole, Robinson, Ansaldo, Whiteman, & Spradlin, 2012; Whiteman, Shi, &
Plucker, 2011).

**Research Questions**

This study focused on investigating the impact of a value-added teacher evaluation model
on teachers in a large urban Indianapolis school district. This research was conducted through
five distinct investigations:

- **VETCS:** Is VAMIN resulting in better student achievement when compared to the
  previous year’s non-VAMIN teacher evaluation?
- **VGCS:** Is VAMIN resulting in better student growth when compared to the previous
  year’s non-VAMIN teacher evaluation?
- **VMRS:** Are there relationships between student achievement growth under a VAMIN-
  based teacher evaluation when controlling for student gender, race, special education
  status, socioeconomic status, Section 504 status, English learner status, and student
  previous year achievement score?
- **VCCS:** Is there consistency between teacher effectiveness ratings and principal-based
  observation evaluations?
- **VSTI:** What impacts do teachers perceive that VAMIN has on teachers in VAMSD?
Results

The following is a brief description of the findings of each investigation.

**VETCS.** VAMIN is not resulting in better student achievement VAMSD. Student mean test scores did not consistently improve at a statistically significant level under VAMIN when compared to non-VAMIN test scores. These results, from a single district, do not support the claims of Demie (2003) and Aaronson, Barrow, and Sander’s (2007) that using VAM’s would increase student achievement. While their arguments as to how the use of VAM’s would improve student results, it is clear that those results did not materialize in VAMSD. This could be because VAM’s are not responsible for improving student performance and teacher self-reflection is the real cause for improvement, (Demie, 2003) or that there are not many effective teachers in VAMSD. Potentially, the reason student achievement did not largely improve is that student achievement is not used to evaluate teachers under VAMIN. Student growth in test score performance year-to-year is used instead.

**VGCS.** VAMIN is consistently resulting in growth in student achievement. Student mean test scores did consistently improve at a statistically significant level under VAMIN when compared to non-VAMIN test scores.

This result is very positive because it shows a benefit of VAM-based teacher evaluations. Additionally, as Ballou (2002) suggested, this is a fairer way to evaluate teachers over student achievement because growth acts to level the playing field and makes teachers accountable for learning gains in all students.

However, gaps exist in the data as students are only tested year-to-year in math and ELA in grades three through eight creating a dual system of accountability. Teachers who teach math and ELA in grades four through eight have TER measures calculated by the Indiana Department
of Education (IDOE) as these are the only teachers that have year-to-year state-mandated standardized assessments. Other teachers use locally designed or selected pre and post-tests to determine student growth. These students are only compared against students within the district unlike TER measures that are calculated against student results across the state. Thus, there is a lack of consistency in terms of calculating student growth percentiles based on subject and grade levels taught. This supports the previous claims of Misco (2008). This means that growth can only be measured in fourth and eighth grades in math and ELA. These results are calculated by the Indiana Department of Education and are proprietary meaning that school personnel and individual teachers cannot check the accuracy of the results. This is because each teacher’s results are based on the results of other students with the same previous-year starting score. Thus, under the Family Educational Rights and Privacy Act (FERPA), student results to their non-school teachers and administrators cannot be disclosed. This lack of transparency as suggested in other models (Kupermintz, 2003) should be addressed by legislators immediately. While growth measures are calculated for other non-math or ELA subjects in grades K-12 locally, this system sets up a dual accountability system, one calculated by the Indiana Department of Education and one calculated by local districts. This could result in future legal challenges.

**VMRS.** Student achievement growth is most impacted by student previous year achievement score with a negative association. This means that as a student’s previous year assessment score increases, the student’s scale score growth decreases. Additionally, VAMIN does not have a consistently strong impact on student achievement growth.

VAMIN does not have the largest standardized coefficient in magnitude. In all models except for three (70 percent), prior-year test score has the largest standardized coefficient in
magnitude which could give an advantage to a teacher teaching low level students under a 
VAMIN framework. This means that out of all the variables analyzed in the multiple regression 
equations, prior test score routinely had the greatest impact on student growth. Additionally, the 
coefficient is negative. This means that students who started with low prior-year scores were 
likely to grow more than a student with a high prior-year score. This idea identified in other 
models by Koedel and Betts (2008) gives an advantage to a teacher teaching low level students 
under a VAMIN framework. This contradicts claims made by Baker et al. (2010) and 
Anderman, Anderman, Yough, and Gimbert (2010) that VAM’s result in perverse incentives in 
which teachers are not incentivized to take on the most troubling students. In VAMIN, teachers 
are more likely to result in better student growth by teaching students with lower previous-year 
scores. This likely has to do with ceiling effects that are present in ISTEP+ tests that limit the 
score a student can obtain. An outstanding student can only demonstrate mastery to the upper-
limit score. Thus, if a student starts at the upper limit and finishes at the upper limit, it appears 
that the student has not made any academic growth. Instead, that student did not have the 
opportunity to show growth. Hence, the ability to demonstrate student growth which is a proxy 
for teacher effectiveness is constrained by the ceiling of the test.

**VCCS.** There is very poor consistency between teacher effectiveness ratings (TER) and 
principal-based observation evaluations. This is an indication that the evaluations are unreliable 
in at least one of the measures if not both as they are weakly correlated and inconsistent as van 
de Grift (2009) has previously suggested.

Also, present, at least partially, is the misidentification of teacher effectiveness in which 
teacher effectiveness changes based on measures used and the timeframe measured as suggested 
in previous research (Baker, et al., 2010; Boardman & Murnane, 1979; Goldhaber & Hansen,
Since the measures weakly correlate, one if not both of these measures could be misidentifying teacher evaluation categorization. This is likely because, as previous analysis indicated, there are other factors that influence student test results outside of teacher impacts such as previous test scores, socioeconomic status, etc. Additionally, principals may value other components of education that are not identified in TER calculations such as empathy, kindness, and dedication (Harris, Ingle, & Rutledge, 2014).

**VSTI.** Teachers generally perceive impacts from VAM-based teacher evaluations as negative. They report high levels of pressure to raise student test scores while simultaneously being demoralized to the point of considering leaving the teaching profession. VAMSD teachers do not believe that student growth measures should be used for teacher evaluations. VAMSD teachers believe that student/teacher cheating has become increasingly concerning through the adoption of VAM-based teacher evaluations, and they believe that VAM-based teacher evaluations have resulted in a reduction in teacher collaboration.

**Conclusions**

Through the research conducted in this study many important conclusions have been brought to light.

**VAM-based Teacher Evaluations Should be Reconsidered.**

First, in the same way other researchers have raised concerns about the use of VAM (Bracey, 2004; Kupermintz, 2003), VAMIN should be reconsidered in VAMSD. VAMIN is causing a great deal of apprehension among teachers and many teachers are considering leaving the teaching profession due to the new evaluation system. Unfortunately, the number of teachers who are leaving the teaching profession due to the new evaluation system is unknown.
Additionally, the current VAMIN evaluation does not count VAM components as a high percentage of a teacher’s evaluation. The VAM-based components in terms of TER do not correlate with a teacher’s summative evaluation. Thus, VAM-based teacher evaluations are creating an inordinate amount of teacher stress, but they are minimally informing teacher evaluations. Also, VAM-based teacher evaluations did not raise student achievement and while student growth in achievement scores increased, the growth was not attributable to the use of VAMIN. While VAM in Indiana is intended to be used in evaluating teachers, VAM and more specifically SGPs were never intended to be used to make high stakes decisions about teachers (Linn, 2008).

As a final point, it is highly likely and suddenly fashionable to suggest that since VAM-based teacher evaluations have tremendous flaws that other elements of evaluation should be added to VAM models such as peer observations, student and parent surveys, portfolios, and teacher self-reports and reflections. Likely, none of these changes will improve VAM-based teacher evaluations because of concurrent validity. This is a measure of correlation between multiple measures to demonstrate whether or not different measures yield similar results. Since these measures will likely not correlate, the VAM-based components will still result in invalid models (Chester, 2003; Gordon et al., 2006; Papay, 2011; Schochet & Chiang, 2010).

**Indiana School Districts Should Increase Support for Teacher Understanding of VAM**

VAMSD teachers have clearly been demoralized by VAM-based teacher evaluations. While it is unlikely that schools will discontinue VAM use, enhanced efforts to ensure teacher understanding of VAM’s strengths and weaknesses should be conducted. Teachers reported that they largely received information about VAM from fellow teachers or building administrators. If
teachers are only getting information from these sources, the information they receive could be inaccurate. This could be causing confusion and stress for teachers.

**Teachers Should Not be Held Solely Responsible for Student Learning**

While the use of VAM-based teacher evaluations might be perceived by non-educators as a step in the right direction toward improving teacher evaluation, it clearly misses the mark in VAMSD. Through this model only student test scores and evaluation perceptions are used to evaluate teachers. Additional controls could and should be included to more thoroughly evaluate teachers (Baker et al., 2013; Ehlert et al., 2012; Goldschmidt et al., 2012; Reckase, 2004). Without these controls there is a high likelihood of lawsuits as teachers are being held accountable for student attributes over which they have no control.

**A Better Framework for Education Should be Employed**

Becker’s (1964) Human Capital Theory along with the Generalized Cumulative Effects Model (CEM) is flawed in that these models view the teacher as a producer. The teaching profession is solely responsible for creating economic inputs for the nation in the form of students educated as future employees. The theory represents future economic opportunity as dependent upon the abilities of the workforce that are derived directly from educational success. Teacher’s jobs are and should be more than just increasing test scores to demonstrate career readiness. Through VAM-based teacher evaluations, teachers are solely responsible for increasing student achievement. However, there are elements of education that are as, if not more important, than student achievement. Thus, there should be a shift away from economics-based educational theories to something more beneficial for students, teachers, and society as a whole.
Implications for Practice

Practitioners

Teachers should continue to voice concerns about the perceived faults of VAM. Teachers should receive greater support and training on the use of VAM. If VAMIN is continued, teachers should seek teaching assignments in classes that are made up of low-scoring students.

Policy Makers

Policy makers should reconsider laws that require teachers to be evaluated by VAM. Instead, VAM-based teacher evaluations should be repurposed as a formative tool to help teachers review and improve their practice. This notion has been supported by the American Statistical Association (ASA) recently (American Statistical Association, 2014). The ASA which is a non-education association comprised of expert statisticians from a multitude of disciplines has been very vocal in stating that VAM-based evaluations should not be used for high-stakes decisions such as job termination and compensation increases.

Theory

Economists should discontinue promotion of theories that attribute student learning solely to teachers such as Human Capital Theory. These theories treat students not as human beings but as economic inputs in a global economic model. Teaching and learning is an incredibly complicated partnership between a multitude of factions including students, teachers, administrators, parents, family members, community members, and politicians. The continued focus on the evaluation of one component is overly-simplistic and often leaves out important factors in the educational process. Additionally, these models treat the sole purpose of the American education system as economic opportunism. This may be a component, but it should
not be what the American educational system aspires to. Our educational aspirations should be much greater and focused on improving the lives of all Americans, socially, emotionally, culturally, and economically.

**Recommendations for Future Research**

The following are recommendations for future research.

**Replication Studies**

Replicate this research on other school corporations to determine whether or not these results occur in other districts with different size and demographics.

**Departed Educator Interviews**

Perform interviews of teachers who have left VAMSD within the last two years to determine why they left the district and if VAMIN played a role in their decision to leave.

**Comparison to Enhanced VAM**

Analyze an enhanced TER that controls for factors that are out of the teachers control such as race, socioeconomic status, special education status, etc. to determine whether the results are different from current TER measures.
REFERENCES


Cole, C. M., Robinson, J. N., Ansaldo, J., Whiteman, R. S., Spradlin, T. E., & Indiana University, Center for Evaluation and Education Policy. (2012). *Overhauling Indiana teacher evaluation systems: Examining planning and implementation issues of school districts*. *Education Policy Brief, Volume 10, Number 4, Summer 2012*. Center for Evaluation and Education Policy. 1900 East Tenth Street, Bloomington, IN 47406-7512. Tel: 800-511-6575; Tel: 812-855-4438; Fax: 812-856-5890; e-mail: ceep@indiana.edu; Web site: [http://www.indiana.edu/~ceep](http://www.indiana.edu/~ceep).


Farmer, B. (2009). VU researcher questions direction of race to the top. *WPLN NEWS.*


Indiana Code 20-28-2-6
Indiana Code 20-28-7.5-1

Indiana Code 20-28-11.5


Reckase, M. D. (2004). The real world is more complicated than we would like. Journal of Educational and Behavioral Statistics, 29(1), 117-120.


Whiteman, R. S., Shi, D., Plucker, J. A., & Indiana University, Center for Evaluation and Education Policy. (2011). *Revamping the teacher evaluation process. Education policy brief. Volume 9, Number 4, Fall 2011.* Center for Evaluation and Education Policy. 1900 East Tenth Street, Bloomington, IN 47406-7512. Tel: 800-511-6575; Tel: 812-855-4438; Fax: 812-856-5890; e-mail: ceep@indiana.edu; Web site: http://www.indiana.edu/~ceep.


APPENDIX A: SURVEY RECRUITMENT LETTER

Subject: Survey: Teacher Perceptions on the Use of Value-Added Measures in Teacher Evaluation Models

Dear Teacher:

My name is Chad E. Michalek, and I am an educational administration doctoral student at Ball State University. I am interested in conducting quantitative research on teacher perceptions towards the use of value-added measures (VAM) in teacher evaluation models.

In the fall of 2012, Indiana schools, your school district included (referred to as VAMSD to ensure anonymity), began transitioning to this type of model that evaluates teachers partially on VAM. The VAM components of your district’s teacher evaluation model (referred to as VAMIN to ensure anonymity) are ratings based on student test scores such as teacher effectiveness ratings (TER) and student growth percentiles (SGP). These measures are a collection of statistical formulas and techniques that try to derive teacher effects on student performance. This project will be comprised of only teachers in your school district. Any information you provide in this survey will be completely anonymous. Your responses may be shared in the aggregate, but your specific responses and identity will be concealed. My intent is not to evaluate you or your responses, only to try and better understand teacher perceptions of VAM-based teacher evaluation models. This survey has been approved by: Superintendent, Dr. XXX XXXXX; VTEA President, Ms. XXX XXXXX; and the Ball State Institutional Review Board (IRB). The survey should take approximately 10 minutes to complete.

The purpose of this research is to determine what fears, reluctances, hopes, and aspirations, if any, teachers feel about the utilization of this evaluation technique. This will be done in an effort to determine if there are any possible areas of improvement that would allow value-added teacher evaluation models to more effectively evaluate teacher performance. The results of this project have the potential to influence future educational policy, improve teacher-administrator relationships, and improve student educational outcomes.

I urge you to consider being a part of this valuable teacher-based research. If you would like to participate, please click the link at the bottom of this email that will connect you to an online survey that asks a variety of questions. If you choose to participate, please complete the survey. Thank you for your consideration in being a part of this project.

If you have any questions or concerns about completing the survey or about participating in this study, you may contact me at 765.729.4259 or at cemichalek@bsu.edu. If you have any questions about your rights as a research subject, you may contact the Ball State University
Institutional Review Board (IRB) Director, Office of Research Integrity, Ball State University, Muncie, IN 47306, 765.285.5070 or irb@bsu.edu.

Sincerely,

Chad E. Michalek
Ball State Doctoral Student
SURVEY LINK
APPENDIX B: TEACHER PERCEPTIONS OF THE USE OF VALUE-ADDED MEASURES IN TEACHER EVALUATIONS SURVEY

Informed Consent Form

You volunteer to participate in a research project conducted by Mr. Chad E. Michalek from Ball State University.

Your participation in this project is voluntary. You understand that you will not be paid for your participation. You may withdraw and discontinue participation at any time without penalty. If you decline to participate or withdraw from the study, no one will know unless you tell them.

Participation involves completing an online survey. Teachers who do not currently teach in VAMSD will be excluded. The survey will take approximately 10 minutes to complete. Responses will be recorded via the statistical program Qualtrics. If you do not want to have your answers recorded, you will not be able to participate.

You understand that the researcher will not identify you by name in any reports using information obtained from the survey, and that your anonymity as a participant in this study will remain secure. Subsequent uses of records and data will be subject to standard data use policies which protect the anonymity of individuals and institutions. The data will be stored online via a secure server that is password encrypted for a period of up to one year beginning May 1, 2014.

There are no known risks if you decide to participate in this study. However, if you feel you need counseling due to your participation in the study you can contact counseling services in your area.

You understand that this research study, IRB protocol 592337-1: Teacher Perceptions on the Use of Value-Added Measures in Teacher Evaluation Models, has been reviewed and approved by the Institutional Review Board (IRB) for Studies Involving Human Subjects at Ball State University. For research problems or questions regarding participants, the Institutional Review Board may be contacted at Director, Office of Research Integrity, Ball State University, Muncie, IN 47306; phone number 765.285.5070, or email irb@bsu.edu. You may also contact the Principal Investigator, Chad E. Michalek at 228 Hidden Glen Drive, Greenfield, IN 46140; phone number 765.729.4259; or email cmichalek@bsu.edu. You may also contact the Faculty Adviser, Dr. Serena Sahloul at Department of Educational Leadership, Teachers College (TC), Room 903, Ball State University, Muncie, IN 47306; phone number 765.285.8413; or email sjsahloul@bsu.edu.

You have read and understand the explanation provided to you. You have had all your questions answered to your satisfaction, and you voluntarily agree to participate in this study.
You have been given the option to print this consent form

I have read, understood, and printed a copy of, the above consent form and desire of my own free will to participate in this study. (If you do not wish to participate in the survey, please close your browser window now.)

- Yes

What is your highest level of degree attainment?

- High School
- Bachelors
- Masters
- Specialist
- Doctorate

What curricular area do you currently teach? (Select all that apply)

- English Language Arts
- Math
- Other

How many full years of experience do you have as an Indiana public school teacher?

- 0
- 1
- 2
- 3 - 5
- 6 - 10
- More than 10
What grade level do you currently teach?

- K - 3
- 4 - 5
- 6 - 8
- 9 - 12

What is your gender?

- Male
- Female

What is your age?

- 21 - 30
- 31 - 40
- 41 - 50
- 51 - 60
- 61 - 70
- Older than 70

What is your race?

- White
- Black
- Asian
- Hispanic
- Multi-racial
- Other
Have you ever worked full-time outside of education?

- Yes
- No

Rate the following individual's levels of statistical expertise.

<table>
<thead>
<tr>
<th></th>
<th>Very low</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
<th>Very high</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Your evaluator</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comment/Clarification
For each statement below, please indicate your level of agreement.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither Agree nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I believe that teacher effectiveness affects student achievement.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe that student achievement can be validly measured.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe that student academic growth can be validly measured.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe that teacher evaluation should be linked to student growth.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe that instruction can be accurately evaluated and judged.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I believe that the relationship between teaching and learning can be accurately applied to an evaluation of teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior to the use of VAM, the teacher evaluation processes in Indiana needed improvement.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An effective teacher evaluation system informs professional development.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An effective teacher evaluation system drives professional development.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAM-based teacher evaluations have resulted in improved teaching and learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher evaluation should be tied to compensation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How familiar are you with Public Law 90 that required the implementation of a new teacher evaluation system starting in the 2012-13 school year?

- Extremely familiar
- Adequately familiar
- Somewhat familiar
- Slightly familiar
- Not at all familiar
Comment/Clarification

How have you become familiar with the requirements of Public Law 90? (Select all that apply)

☐ Read the legislation
☐ Attended workshops/seminars
☐ Participated in webinars
☐ Had discussions with administrators
☐ Discussed requirements of the law with other teachers
☐ Spoke with IDOE officials or reviewed information on the IDOE website

Comment/Clarification

Which stakeholder groups that you know of have been or will be a part of your district's VAM-based teacher evaluation development process? (Select all that apply)

☐ Parents
☐ Students
☐ Teachers
☐ Teacher association leadership
☐ Principals
☐ Central office staff
☐ Technology personnel
☐ Community members
☐ Others
<table>
<thead>
<tr>
<th>Item</th>
<th>Not at all concerned</th>
<th>Slightly concerned</th>
<th>Somewhat concerned</th>
<th>Very concerned</th>
<th>Extremely concerned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resources to conduct classroom observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resources to collect student performance data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resources to provide training for evaluators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resources to provide training for staff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resources for increased compensation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building the capacity for understanding among school personnel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication to key stakeholders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-going support for professional development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear guidance concerning the interpretation of new law</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alignment of new law with policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comment/Clarification

Have you ever considered leaving the teaching profession?

- Yes
- No

Comment/Clarification
Have you considered leaving the teaching profession specifically because of the new evaluation system?

- Yes
- No

Comment/Clarification

Would you recommend the teaching profession to prospective employees?

- Yes
- No

Comment/Clarification

Do you feel pressured to raise student test scores?

- Yes
- No

Comment/Clarification

Has your willingness to help fellow teachers decreased due to the new teacher evaluation?

- Yes
- No

Comment/Clarification
Have you ever witnessed cheating on high stakes testing?

- Yes
- No

Comment/Clarification

Is student/teacher cheating of greater concern to you due to the new teacher evaluation?

- Yes
- No

Comment/Clarification
APPENDIX C: ISTEP+ CUT SCORES

### ISTEP+ Cut Scores

<table>
<thead>
<tr>
<th>Grade</th>
<th>English / Language Arts Pass</th>
<th>Pass +</th>
<th>Mathematics Pass</th>
<th>Pass</th>
<th>Science Pass</th>
<th>Pass</th>
<th>Pass +</th>
<th>Social Studies Pass</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>417</td>
<td>521</td>
<td>413</td>
<td>513</td>
<td>463</td>
<td>535</td>
<td>483</td>
<td>550</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>437</td>
<td>535</td>
<td>445</td>
<td>541</td>
<td>463</td>
<td>535</td>
<td>488</td>
<td>549</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>468</td>
<td>548</td>
<td>463</td>
<td>556</td>
<td>483</td>
<td>550</td>
<td>486</td>
<td>545</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>478</td>
<td>579</td>
<td>487</td>
<td>590</td>
<td>488</td>
<td>549</td>
<td>486</td>
<td>545</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>501</td>
<td>584</td>
<td>511</td>
<td>603</td>
<td>488</td>
<td>549</td>
<td>486</td>
<td>545</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>508</td>
<td>627</td>
<td>537</td>
<td>641</td>
<td>488</td>
<td>549</td>
<td>486</td>
<td>545</td>
<td></td>
</tr>
</tbody>
</table>

### ISTEP+ Minimum & Maximum Scores

<table>
<thead>
<tr>
<th>Grade</th>
<th>English / Language Arts Minimum</th>
<th>Maximum</th>
<th>Mathematics Minimum</th>
<th>Maximum</th>
<th>Science Minimum</th>
<th>Maximum</th>
<th>Social Studies Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>120</td>
<td>780</td>
<td>140</td>
<td>735</td>
<td>200</td>
<td>750</td>
<td>220</td>
<td>760</td>
</tr>
<tr>
<td>4</td>
<td>140</td>
<td>800</td>
<td>185</td>
<td>750</td>
<td>200</td>
<td>765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>160</td>
<td>820</td>
<td>245</td>
<td>775</td>
<td>200</td>
<td>765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>180</td>
<td>850</td>
<td>290</td>
<td>790</td>
<td>200</td>
<td>765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>210</td>
<td>870</td>
<td>315</td>
<td>810</td>
<td>200</td>
<td>765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>230</td>
<td>890</td>
<td>340</td>
<td>830</td>
<td>200</td>
<td>765</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX D: ADDITIONAL TABLES OF SURVEY PARTICIPANT DEMOGRAPHICS

**Table A.1**

*Survey Question: What is your highest level of degree attainment?*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High School</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>Bachelors</td>
<td>30</td>
<td>36%</td>
</tr>
<tr>
<td>3</td>
<td>Masters</td>
<td>45</td>
<td>54%</td>
</tr>
<tr>
<td>4</td>
<td>Specialist</td>
<td>4</td>
<td>5%</td>
</tr>
<tr>
<td>5</td>
<td>Doctorate</td>
<td>5</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>84</td>
<td>100%</td>
</tr>
</tbody>
</table>

**Table A.2**

*Survey Question: What curricular area do you currently teach? (Select all that apply)*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>English Language Arts</td>
<td>30</td>
<td>36%</td>
</tr>
<tr>
<td>2</td>
<td>Math</td>
<td>24</td>
<td>29%</td>
</tr>
<tr>
<td>3</td>
<td>Other</td>
<td>50</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>84</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Table A.3**

*Survey Question: How many full years of experience do you have as an Indiana public school teacher?*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>5</td>
<td>6%</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>9</td>
<td>11%</td>
</tr>
<tr>
<td>4</td>
<td>3 - 5</td>
<td>11</td>
<td>13%</td>
</tr>
<tr>
<td>5</td>
<td>6 - 10</td>
<td>20</td>
<td>24%</td>
</tr>
<tr>
<td>6</td>
<td>More than 10</td>
<td>38</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>84</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table A.4

Survey Question: What grade level do you currently teach?

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>K - 3</td>
<td>11</td>
<td>13%</td>
</tr>
<tr>
<td>2</td>
<td>4 - 5</td>
<td>11</td>
<td>13%</td>
</tr>
<tr>
<td>3</td>
<td>6 - 8</td>
<td>28</td>
<td>34%</td>
</tr>
<tr>
<td>4</td>
<td>9 - 12</td>
<td>33</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>83</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table A.5

Survey Question: What is your gender?

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>32</td>
<td>39%</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>51</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>83</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table A.6

Survey Question: What is your age?

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21 - 30</td>
<td>23</td>
<td>27%</td>
</tr>
<tr>
<td>2</td>
<td>31 - 40</td>
<td>21</td>
<td>25%</td>
</tr>
<tr>
<td>3</td>
<td>41 - 50</td>
<td>14</td>
<td>17%</td>
</tr>
<tr>
<td>4</td>
<td>51 - 60</td>
<td>19</td>
<td>23%</td>
</tr>
<tr>
<td>5</td>
<td>61 - 70</td>
<td>7</td>
<td>8%</td>
</tr>
<tr>
<td>6</td>
<td>Older than 70</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>84</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table A.7

Survey Question: What is your race?

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>White</td>
<td>75</td>
<td>89%</td>
</tr>
<tr>
<td>2</td>
<td>Black</td>
<td>5</td>
<td>6%</td>
</tr>
<tr>
<td>3</td>
<td>Asian</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>Hispanic</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>5</td>
<td>Multi-racial</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>6</td>
<td>Other</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>Total Responses</td>
<td>84</td>
<td>100%</td>
</tr>
<tr>
<td>#</td>
<td>Answer</td>
<td>Response</td>
<td>%</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
<td>32</td>
<td>38%</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>51</td>
<td>52%</td>
</tr>
<tr>
<td>Total Responses</td>
<td>84</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>