# TEACHER MERIT PAY IN A RURAL WESTERN NORTH CAROLINA COUNTY: A QUANTITATIVE ANALYSIS OF THE EFFECTS OF STUDENT CHARACTERISTICS ON A TEACHER'S LIKELIHOOD OF RECEIVING A MONETARY BONUS IN MATH OR READING IN GRADES THREE - EIGHT

A Dissertation
by
ROSS RUMBAUGH

Submitted to the Graduate School
at Appalachian State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF EDUCATION

December 2019
Educational Leadership Doctoral Program
Reich College of Education

TEACHER MERIT PAY IN A RURAL WESTERN NORTH CAROLINA COUNTY: A QUANTITATIVE ANALYSIS OF THE EFFECTS OF STUDENT CHARACTERISTICS ON A TEACHER'S LIKELIHOOD OF RECEIVING A MONETARY BONUS IN MATH OR READING IN GRADES THREE - EIGHT

A Dissertation
by
ROSS RUMBAUGH
December 2019

APPROVED BY:

______________________________________
Les Bolt, Ph.D.
Chairperson, Dissertation Committee

______________________________________
William M. Gummerson, Ph.D.
Member, Dissertation Committee

______________________________________
Vachel Miller, Ed.D.
Member, Dissertation Committee

______________________________________
Vachel Miller, Ed.D.
Director, Educational Leadership Doctoral Program

______________________________________
Mike McKenzie, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

## Abstract


TEACHER MERIT PAY IN A RURAL WESTERN NORTH CAROLINA COUNTY: A QUANTITATIVE ANALYSIS OF THE EFFECTS OF STUDENT CHARACTERISTICS ON A TEACHER'S LIKELIHOOD OF RECEIVING A MONETARY BONUS IN MATH OR READING IN GRADES THREE - EIGHT

Ross Rumbaugh
B.A. with Honors in Economics, Wake Forest University
Teaching Licensure, Winston-Salem State University
MSA, Appalachian State University
Ed.S., Appalachian State University
Ed.D., Appalachian State University

Dissertation Committee Chairperson: Les Bolt, Ph.D.

This quantitative work is an exploratory study that examines the bonus pay structure enacted by the North Carolina General Assembly in the 2016-2017 academic year for teachers in grades three through eight in reading and math. The study used student data collected from an LEA in a rural western North Carolina county from the 2017-2018 academic year. The study and logistic regression analyze the validity of the EVAAS value-added tool used in North Carolina to identify those educators that rose to the top 25% of teachers in the state in their ability to impact student learning in the affected grade levels and subject areas. With the non-random assignment of students into elementary and middle school classrooms and the corresponding peer effects that either lift or retard student learning, the study examined student characteristics to find any correlation between a teacher receiving

the merit-based bonus and the composition of students in her classroom. The study identified that white, mixed-race and Asian students have a greater likelihood of sitting in a teacher's classroom that received the reading bonus. Students with disabilities had a negative correlation to a teacher's likelihood of receiving the reading bonus in grades three - five. In math, the study found a negative correlation between teachers receiving the bonus and the placement of both students with disabilities and students labeled as gifted learners in their classrooms. There was a positive correlation between teachers receiving the math bonus and students that are mixed-race, Hispanic and African-American. The study might be used to help inform student classroom assignment practices in North Carolina in light of a current merit-based pay legislation that is rewarding teachers for the growth they achieve with the students they serve.

**Acknowledgements**

I would like to express my gratitude to my dissertation team for working with me for too many years on this project. Dr. Les Bolt has hung in there with me through topic changes and rewrites that might have encouraged less patient people to excuse themselves from my dissertation committee. Dr. William Gummerson's tireless editing of my work has been greatly appreciated. His attention to detail was invaluable to my writing process. And finally, Dr. Vachel Miller came to the rescue when I found myself down a committee member at the eleventh hour. Thank you to all three men for their dedication to my work.

I would like to thank Sally McIlwain for the hours she spent collecting and collating the data so that it maintained for me a level of anonymity while conducting my research. Thank you also to Brittany White and Dr. Melanie Honeycutt for their formatting expertise. Thank you to Dr. Kristin Edwards, a recent dissertation defender herself, for her reviews and recommendations for my research. I want to thank Brent Baker for conducting a *Statistics 101* review with me as I began the research process. Rhonda Rumbaugh was my family-level cheerleader, asking me every holiday season, in detail, the progress of my work. She listened intently to each title change and topic iteration over too many seasons. In addition, I want to give a special level of thanks to Dr. Larry Putnam who would not let me quit when I was on the verge of throwing in the towel. Thank you for believing in my process and motivating me to complete the journey when it was looking like an insurmountable task.

## Dedication

Although he passed away years ago, my grandfather, Dr. Ross St. Clair Rumbaugh, was the first Rumbaugh doctor. A small town surgeon and family practice doctor, he was an inspiration for my pursuing the masters, Ed.S, and finally doctoral degree. I loved the sound of his baritone voice that would impart to me words of wisdom, hoping that I may learn the value of discernment in making education, career, and family decisions. Sitting beside me at the table during these conversations were my father (also Ross), a 32 year veteran of the public school classroom, and my mother Sharon, the most selfless person I've ever met and the reason I've had any personal and professional success in my lifetime. The after-Thanksgiving dinner discussions were the greatest part of the cold Pennsylvania holiday season.

I want to dedicate this work also to my daughter Ellie and my son, Brittain. God gifted you both with athletic, musical, and academic talents. May you use these talents wisely to further His kingdom. I hope to see you both reach whatever educational aims you desire in order to attain the career, family, and faith goals you set for yourselves.

And finally to my wife, Marissa, thank you for putting up with my never ending educational pursuits, my career, and my moods. You are a wonderful wife, mother, and supporter. I love you and thank you for being all of those things to our family.

## Table of Contents

## List of Tables

## Chapter 1: Introduction

Teacher evaluation is not a perfect science. Although researchers have found that the teacher is the most important school-based variable affecting student learning, administrators have struggled to evaluate teachers in a way that discriminates between the most and least effective (Marzano, Pickering & Pollock, 2001; Weisberg, Sexton, Mulhern & Keeling, 2009; Sanders, Wright, & Horn, 1997). In most school districts, teacher evaluation is a notoriously subjective exercise that is rarely directly linked to student achievement (Braun, 2005). Done correctly, teacher evaluation should be used to identify the most effective teachers for continued employment, and to discover areas where professional development is needed. The most accurate teacher observations appear to be those conducted by an outside agent who is not biased by day-to-day contact with the teacher (Whitehurst, Chingos, & Lindquist, 2015). However, smaller school districts and those in poorer communities often cannot afford the personnel to complete these outside observations. The challenge then for public school administrators, district school offices, and state departments of education is to find some level of objective measurement concerning what constitutes effective teaching practices. And of equal importance is the ability to utilize this measurement tool to identify effective and ineffective teachers for purposes of merit pay, teacher retention, staff development, and teacher dismissal (Weisberg et al., 2009).

The history of education reform resembles a pendulum that swings between absolute autonomy of the classroom teacher to over-prescribed instructional methods derived in part from the scientific management practices advocated by Frederick Taylor in the early 20th century. Taylor believed that if managers simplify tasks into their smallest components, overall production improves (Ireh, 2016). Applied to academia, in order to maximize production (student results), the outputs must be measurable so that the efficient workers (teachers) can be recognized and rewarded, the least effective dismissed (Gray, 1993). Professionals in the educational field believe that the results of student assessment can be used to determine whether students have mastered the curriculum, and their teachers have been effective at their craft.

A growing debate exists, however, concerning what student testing reveals about teacher effectiveness. On one side are those that want to assign accountability to teachers as measured by the growth of their students on End of Grade (EOG) and End of Course (EOC) tests (Ballou, Sanders, & Wright, 2004; *Race to the Top*, 2011). On the other side of the debate are those who worry about the validity of these accountability tools and the historical tendency of state and federal legislative mandates to over-regulate classroom teaching (Amrein-Beardsley & Collins, 2010; Collins & Amrein-Beardsley, 2014; Darling-Hammond, Amrein-Beardsley, & Rothstein, 2012).

When it comes to using standardized test scores to hold public school teachers accountable, several questions arise. What kind of measurements of teacher performance should the public expect? If teacher observations and evaluations are subjective in nature, what measurement tools can education leaders incorporate in the evaluation process that will insure that students are gaining what they need year by year to be successful in a civil society (Whitehurst, Chingos, & Lindquist, 2015)? Are the tools that are currently being used to measure student academic growth and teacher impact sufficiently valid and reliable to meet their intended purpose (Amrein-Beardsley, 2014)?

**Competing Paradigms**

The search for a reliable tool for measuring teacher effectiveness is laudable. Finding the right one, or even one that accurately reflects teacher effectiveness based on student test scores is not easy, and may ultimately be impossible. Of the Value-Added-Measures (VAMs) on the market, Wiley (2006) discusses six that he alternately refers to as Value-Added-Assessments (VAAs) being used by states and districts across the country.

- Simple *gain score* models (e.g., Texas Growth Index)
- Covariate adjustment models (e.g., Dallas Value Added Assessment System or "DVAAS")
- William Sanders's layered models (e.g., Educational Value Added Assessment System or "EVAAS"; Tennessee Value Added Assessment System or "TVAAS")
- Cross-classified models
- Rand's persistence model
- Todd and Wolpin's cumulative within-child mixed-effects model (p. 3).

North Carolina adopted the Sanders' EVAAS model in 2001 to measure the impact teachers have on student learning, often referred to simply as *growth*. For the purposes of discussion going forward, *growth* is the amount of academic progress a student makes in a year of study (Cody, McFarland, Moore & Preston, 2010). Since its adoption, the North Carolina EVAAS tool has been deluged with a mountain of critical literature that debates the accuracy of any VAM model, especially those that may be used to make high stake decisions about teacher evaluation, compensation and termination (Amrein-Beardsley, 2014). Many value-added-model critics believe that no value-added model or methodology is sufficiently robust to measure teacher effectiveness. These VAM critics purport that no level of mathematical modeling is sophisticated enough to isolate the value a teacher has on the learning of one child or classroom (Baker et al., 2010.)

A contrary opinion on the validity of VAMs arises from various state legislatures and VAM creators that claim to measure accurately student growth. As referenced previously, these models consist of the Education Value-Added Assessment System (EVAAS), the Milwaukee model, and the Dallas model to name a few (Cody et al., 2010) (Appendix A).

> The Dallas and Milwaukee school districts use value-added models that differ from EVAAS in their use of demographic data. In addition to test scores, these two models incorporate information on student characteristics, such as race and income level. In both cases, the researchers who developed these models have specified that the models are to be used for measuring teacher and school effectiveness, not for setting performance standards. (Cody et al., 2010, p.3)

These student growth models use complex statistical techniques to isolate the influence of individual school districts, schools, and teachers on student achievement (Meyer & Dokumaci, 2010).

Used in North Carolina since 2001, EVAAS gages the impact teachers have on improving student outcomes as measured on end of year state assessments (ABCs of Public Education, 2006). In addition to its usage in North Carolina, several other states and municipalities across the country utilize the EVAAS growth measure (known by other acronyms in different states) to measure teacher effectiveness. In the succeeding chapters, I will present arguments in favor of and in opposition to the use of value-added methodology in making high stakes decisions including merit pay, teacher evaluation, and employment decisions, referencing a myriad of studies in the literature available on the validity and reliability of value-added-models.

**Problem Statement**

In a search for objectivity in the process of evaluating teacher effectiveness, states have turned to value-added models to assign growth values to their teachers. According to the North Carolina Department of Public Instruction's (NCDPI) closeout report of 2015, the state has made strides to make possible every educator's ability to receive a "fair, rigorous, data-based evaluation of his or her effectiveness that includes specific feedback to fuel professional growth" (NCDPI, 2015, p. 14). In part by adopting the EVAAS value-added measure, North Carolina received a federal *Race to the Top* (RttT) grant of close to $400 million in 2010 as one of the grant's original 12 recipients (Klein 2014; NCDPI, 2015). As part of their grant application, NCDPI agreed to rate teacher effectiveness on a three-year rolling average that began in 2014-15. This rolling average became Standard 6 in the teacher evaluation tool and was to be used to measure teacher effectiveness over a three-year period in order to achieve a more complete picture of a teacher's effectiveness (Appendix B). According to the *North Carolina Teacher Evaluation Process* (2015) crafted by the North Carolina Department of Public Instruction:

> *Standard VI: Teachers Contribute to the Academic Success of Students*. The work of the teacher results in acceptable, measurable progress for students based on established performance expectations using appropriate data to demonstrate growth. (p. 5)

The data from three years of teacher impact on student growth as measured by EVAAS would designate a teacher as performing at one of three levels: *Highly Effective*,

*Effective*, or *Needs Improvement* (NCDPI, 2015, p. 6). Until 2017, teachers were able to see data on the *North Carolina Educator Effectiveness System* (NCEES) evaluation site and access their three-year effectiveness rating based on the average of their three-year growth indexes as calculated in EVAAS. However, in 2017, the state ended the practice of rating teacher effectiveness using the three-year rolling average. Year by year teacher growth numbers are still housed in the *Educator Evaluation Dashboard* in the EVAAS web platform and can be accessed by the teachers and the administrators for whom these teachers work (NCDPI, 2015).

Although "Standard 6" is no longer a part of the teacher evaluation instrument, a teacher's impact on student growth as measured by EVAAS in now a part of the North Carolina teacher bonus compensation package. For all states trying to apply for *RttT* money, it was required that they provide in their grant applications a methodology that would identify the state's most and least effective teachers. The intent of the federal legislation was to encourage states to identify objectively their most impactful teachers. This is precisely what EVAAS and other VAMs claim to do.

**Teacher Compensation Plans**

With student performance as the goal, North Carolina has taken steps since 2014 to revamp its teacher compensation plan by placing less emphasis on years of service and degrees earned and instead choosing to reward teachers for the growth of their students (NC House Bill 840, 2017). Starting with the 2016-17 school year, the top 25% of public school teachers, as measured by their EVAAS growth index in grades 3-8, began earning incentive bonuses for growing their students. Legislation was passed in the North Carolina General

Assembly that was to motivate teachers and reward those that most impacted student learning. According to Eckert and Dabrowski (2010):

> The current teacher compensation system is obsolete. On that point, there appears to be wide agreement. Virtually every new idea for improving teacher compensation focuses on how to tie teacher pay to teacher effectiveness as measured by improving student learning. (p. 88)

In light of the weight being attributed to the EVAAS growth measure to evaluate teachers and assess student growth, it is reasonable to ask about their accuracy, most especially because of the high stakes decisions being made around teacher compensation, teacher employment and principal placement (Goldring, Grissom, Rubin, Neumerski, Cannata, Drake, & Schuermann, 2015).

Using EVAAS measures to make high stakes decisions about teacher compensation and retention was not the intent of value-added modeling at its inception. Rather, Sanders and Horn reported in 1998 that the purpose of the EVAAS tool was not to make high stakes decisions, but to allow teachers the ability to reflect and improve on their teaching.

A qualitative study conducted by Collins and Amrein-Beardsley (2012) revealed however, that teachers have an inherent distrust of the EVAAS instrument due to the VAMs lack of transparency. Can EVAAS results mitigate the "peer effect" that other researchers have found to lead to higher student scores or lower student scores independent of the classroom teacher (Burke & Sass, 2013; Hanushek, Kain, Markman, & Rivkin, 2003; Sacerdote, 2011)? Can EVAAS account for "spillage*,*" the phenomena of a teacher's effectiveness rating being contaminated by the influence of another teacher or tutor on a student's performance and growth (Corcoran, 2010; Koedel 2009, Yuan, 2015)? Should teachers be blamed for lower student growth when summer learning loss occurs (Darling-Hammond et. al, 2012; Harris, 2010)? Can EVAAS account for the preponderance of English as a second language learners in a classroom or academically gifted students being hindered by the "ceiling effect" with their end of year assessments (Collins & Amrein-Beardsley, 2014; Sanders & Horn, 1998)? Can any VAM model reflect teacher effectiveness accurately when students are not randomly assigned to classrooms (Burns & Mason, 1995; Paufler & Amrein-Beardsley, 2014, Baker et al., 2010)? This study will explore whether there are student characteristics that may inhibit a teacher's ability to achieve higher EVAAS indexes precluding any bonuses the state of North Carolina will award to the top 25% of teachers in grades three through eight (Ballou, Sanders, & Wright, 2004). Are all teachers equally able to earn a merit-based bonus in spite of the differing characteristics of the students in their classrooms (Eckert & Dabrowski, 2010)?

**Purpose**

The focus of this study is to determine whether characteristics of students in a North Carolina teacher's classroom – by subject area and grade – have an effect on that teacher's ability to finish in the top 25% of EVAAS scores. When it comes to educating students, the end should always be an improvement in student performance (Guskey, 2017). It is imperative that North Carolina teachers understand the system that is evaluating their teaching effectiveness and that the instrument used to measure student academic growth is a valid and reliable instrument.

In North Carolina, EVAAS is the value-added-model the state has adopted to measure a teacher's impact on the learning of her students. Put simply, is EVAAS an appropriate tool for measuring teacher effectiveness? Ballou and Springer (2015) found evidence that in states where EVAAS is used, teachers failed to claim students in roster verification appropriately if they were low performers, and found also that teachers were coaching (if not cheating) at a statistically significant level when teachers were in charge of testing their own students. In addition, since EVAAS does not consider student demographic or socio-economic variables in their growth model, one must ask if the model that is calculating teacher growth indexes is a fair measure in light of the fact that demographics and student characteristics are not specifically included in the growth calculations. Are there specific student characteristics that are likely to unduly burden or advantage a teacher's effectiveness measure, resulting in her receiving or not receiving merit pay for student achievement in grades three through five reading and four through eight math? Based on the Guskey framework (2017), do these monetary incentives lead to better student learning outcomes across North Carolina?

The purpose of this study is to test EVAAS' claim that the teacher growth model, despite being void of student demographic and socio-economic considerations, and based solely on individual student testing histories, is sufficiently robust to accurately reflect teacher effectiveness in a set of rural elementary and middle schools in western North Carolina, where students are non-randomly assigned to classrooms and where teacher bonuses are tied to their EVAAS performance.

**Research Question**

In a 2009 article, William Sanders teamed up with Jill G. Leandro, June C. Rivers, and S. Paul Wright, to respond to seven common concerns levied at EVAAS about its ability to reflect accurately teacher effectiveness. The seven concerns the authors chronicled provide context for the two (in bold) that are germane to this discussion and study.

1. Value-added models rely on standardized tests, which have limitations themselves
2. Missing student test data jeopardize the validity of the analyses
3. **Potential for rewards and punishments is related to class size (shrinkage estimation)**
4. **SAS EVAAS does not adjust for socioeconomic factors**
5. SAS EVAAS modeling lacks transparency and is too complex
6. SAS EVAAS statistical methods and algorithms have not been peer reviewed
7. SAS EVAAS predictions of student performance are not verified later (emphasis mine).

McCaffrey, Han and Lockwood (2008) found that the shrinkage estimation that EVAAS and other multivariate models use helps to limit the strong effects that outliers may have on a teacher's results when her class size is small. Sanders et al. (2009) responded to other criticisms about their model not adjusting for socio-economic or demographic variables. Sanders et al. (2009) offered that what the EVAAS model does include is a student's entire testing history in multiple subjects (multivariate) over multiple years (longitudinal) (p.3). At the student level, by including all of his testing history, each student serves as his own control (Sander et al., 2009). Both Ballou et al. (2004) and Lockwood and McCaffrey (2007) confirm Sanders' claims in their research of his EVAAS model. Sanders et al. (2009) include in the defense of their model a thought-provoking statement:

> On a philosophical level, the question educators should ask is whether they should have lower expectations for a student from a poor family than one from a rich family, even when the two students have identical test scores and academic histories. By adjusting for these variables, one is directly assuming that there will be different

> expectations for two students with the same prior achievement pattern who come from different SES/DEM communities. (p. 3)

Sanders and his colleagues admit that there is a strong correlation between socio-economic variables, minority demographics and achievement scores, but they challenge the notion that growth measures have the same negative correlation with these same variables (Sanders et al., 2009, p.4). Sanders et al. (2009) also provide an explanation for why lower EVAAS scores seem to go to teachers in schools serving more impoverished communities. Schools with higher proportions of poor and minority students have been documented to contain fewer seasoned and accomplished teachers, and that strong teachers often leave these schools for higher paying teaching jobs where student poverty is not as prevalent (Mayer, Mullens, & Moore, 2000; Sanders et al., 2009).

However, many researchers find that Sanders underestimates what is going on in these socio-economic areas where low SES students are concentrated in teachers' classrooms. The positive peer effects documented by Hanushek, Kain, Markman, and Rivkin (2003) for students lucky enough to be with the academically gifted student cluster and the teachers who reap the growth rewards are not figured into the EVAAS model, nor are the negative peer effects when clusters of low achieving students populate a classroom (Burke & Sass, 2013; Sacerdote, 2011). Does EVAAS present a model sufficiently complex to handle the lack of random student assignment in most classrooms in America's public schools (Braun, 2005; Kupermintz, Shepard, & Linn, 2001; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Pauffler & Amrein-Beardsley, 2012)? Understanding which of these competing ideas are correct carries significance for North Carolina's public schools. If the stated goal of the North Carolina Legislature is to improve teaching practices and student learning by employing monetary incentives for North Carolina's public school teachers, one must ask if the EVAAS tool is a fair way to determine the most effective teachers in our grades three through eight classrooms, in spite of the fact that teacher rewards are being meted out without consideration to the characteristics of the students sitting in non-randomly assigned classrooms. The research question that emerges from this VAM and student characteristic discussion is as follows: Is the likelihood of a North Carolina teacher receiving incentive pay impacted by the student characteristics prevalent in his/her classroom?

**Setting of the Study**

I will be working with data from 15 elementary schools and six middle schools in a rural western North Carolina school district where close to 62% of the students in grades three through eight in the study are served by the federal *Free and Reduced Lunch* program. The English as a Second Language (ESL) program serves about nine percent of the students. Close to five percent of the students are African American, and another five percent identify as mixed race. Sixty-five percent are white – non-Hispanic. Approximately six percent are Asian and seventeen percent are Hispanic. About seventeen percent of the students are served by the Exceptional Children's (special education) program. Close to seventeen percent are also served by the Academically Gifted Program (Table 3.1).

**Methodology**

The methodology will involve using logistic regression modeling to discover if teacher growth, which determines teachers eligible for state-appropriated merit bonuses, meted out to teachers in grades three through eight in a rural North Carolina public school district, is adversely affected by the demographics and relative poverty of the students in teachers' classrooms. Based on data availability, I will build the regression model with the following variables that are available through the NCDPI Accountability website and the state's secure shell available to district administrators. The student variables I will use to create the regression models include:

- Asian students
- African American students
- Hispanic students
- Native American students
- Mixed race students
- Pacific Islander students
- White/Non-Hispanic students
- Males/Females
- Students served by the Academically Gifted program
- Students in the Exceptional Children's program (special education)
- Economically disadvantaged students identified by their enrollment in the free and reduced lunch program
- English as a second language students

I will identify student characteristics and classroom population clusters that would more likely lead to a teacher receiving or not receiving a bonus for being in the top 25% of EVAAS growth indices.

**Significance**

Currently, 25% of public school teachers across North Carolina teaching math in grades four through eight, and reading in grades three through five, are given annual incentive bonuses upwards of $7000 for effectively growing their students' academic achievement as measured by the value-added measurement tool, EVAAS. Discovering if there is bias inherent in the instrument due to the prevalence of certain student characteristics in teachers' classrooms would be of significance to teachers themselves and those legislators who believe that these bonuses can drive teachers to excel in improving student performance on North Carolina EOG tests. Despite criticisms levied at EVAAS for high stakes decisions based upon its teacher effectiveness results – specifically the effect of non-randomly assigned students and peer effects on EVAAS teacher indices, I have found no research examining the impact of student characteristics on the North Carolina incentives program and the likelihood of teachers achieving a bonus based on their EVAAS results. Furthermore, if policy makers at the state level want to be good stewards of taxpayer dollars, they should know if their teacher bonus plan is having the desired effect of improving student educational outcomes.

**Connections**

Being in education for 25 years in the role of teacher substitute, classroom social studies teacher, basketball, baseball, lacrosse, and cross country coach, assistant principal, principal, and a director of testing and accountability, I have worked with the EVAAS growth tool on several levels. I currently provide trainings to school level administrators, instructional coaches and teachers on the use of EVAAS to enhance instructional practices so that student learning is improved. I have a stake in understanding if the work holds up to the scrutiny of research and to the results generated by this study.

**Definition of Terms**

**Binary Variable –** A categorical variable that can take on exactly two values is termed a binary variable or dichotomous variable.

**Blocking –** When a student's testing history completely accounts for his SES/DEM factors, in effect, blocking out the need to include these variables in a VAM model.

**Categorical Variable** – These are variables that take on one or a limited and usually fixed number of possible values, assigning each unit of observation to a nominal category.

**Ceiling Effect** - The inability of students at the highest achievement levels to show growth due to their previous test scores being at the top of the percentile rankings.

**Descriptive Research Design –** A quantitative research design wherein a subject is measured once and where the researcher is looking only to establish relationships among variables, not to establish causality.

**Experimental Research Design -** A type of quantitative research where subjects are measured before and after a treatment. The researcher is trying to establish causality between variables.

**Exploratory Research** – Research conducted for a problem that must be studied more clearly, intended to establish priorities and improve the final research design.

**Growth** – The amount of gain a student, classroom, or teacher makes in a given year compared to the average growth made by students across the state on a prescribed curriculum

**Logistic regression** - regression analysis used when the dependent variable is dichotomous (binary). **Logistic regression** describes data and explains the relationship between one dependent binary variable and one or more nominal or ordinal independent variables, by computing odds ratios.

**Low Performing School** – In the 2001 *ABC's of North Carolina* model, this designation was for schools that did not meet at least expected growth and had fewer than 50% of its students performing at grade level (achievement level III).

**Merit Pay –** The idea of rewarding teachers with monetary incentives for their impact on student learning rather than for years of service or degrees earned.

**Nominal Variable** – Variables with two or more categories not having any kind of natural order or numeric value.

**Peer Effect –** The effect, positive or negative, that a group of students may have on their fellow students' academic performance.

**Principal-agent theory –** In education, pay for performance runs the risk of the *agent*, the principal in this case, working to maximize student outcomes even though they are not perfectly aware of everything the teacher, the *principal* in this case, is doing to grow their students' academic acumen.

**Scientific Management Theory (Taylorism) -** The application of industrial engineering principles to create a system where waste is avoided, the process and method of production is improved, and goods are fairly distributed.

**Spillage** - A phenomena in education statistics where a teacher gains the advantage or disadvantage of student growth based on the value another teacher adds to that student's learning.

**Status Model** - A model that reports the percentage of students scoring "proficient" on a standardized test. Also known as a proficiency model.

**Student Growth** – The amount of academic progress a student makes in a school year.

**Student Proficiency** – The measure of student performance against a defined level of achievement**.**

**Value-Added-Model** (Measure) - VAMs calculate a measure of change, or the value added to a student's learning over a period of time on the basis of her growth on a standardized test.

## Organization of Study

This study begins with an introduction of the problem, the research question, and the setting and significance of the study involving value-added-models and the North Carolina teacher bonus structure for teachers in grades three through eight. The second chapter examines the literature and informs the study regarding merit pay, EVAAS critiques, EVAAS defenses, the randomization or lack thereof concerning student assignment to classrooms, and the peer effects that influence student learning. Chapter three outlines the explorative quantitative methodology employed in the study. Chapter four examines the results of the quantitative methods and Chapter five provides concluding thoughts about the study and ideas for further research.

## Chapter 2: Literature Review

### Introduction

In 1965, Lyndon Johnson's War on Poverty initiatives paved the way for an educational funding source designed to target early elementary reading. Title I funding became an integral component of the *Elementary and Secondary Education Act* (ESEA) that would be used for early grade reading intervention in an attempt to close the achievement gap between rich and poor (Klein, 2015). Despite the billions of dollars allocated to the Title I program, the educational achievement gap has not been resolved. In 1983, the U.S. Department of Education released a scathing report called a *Nation at Risk.* The rhetoric in this report discredited the nation's public schools and the achievement of its students as compared to the rest of the world. The report found academic rigor lacking in America's public schools. Based on the findings of the report, a *Nation at Risk* facilitated a public conversation about which students should be finishing high school and matriculating to college. According to the report, no longer should universities be reserved for the American elites. College and universities should be accessible to a much larger percentage of the American student population (Guthrie & Ray, 2008).

The political authority that would be responsible for facilitating structural change to the American educational system based on the *Nation at Risk's* findings was the next conversation the nation needed to have. In the mid 1980's, the political authority for operating the nation's public schools was shared by the state and local governments. The state and municipal governments were primarily responsible for funding and directing policy for the public schools within the American system of federalism (Manna, 2006). In this vein, North Carolina was one of the first states to start looking at accountability for its public schools and teachers. To this end, the state began creating state assessments aligned to its curricula to start the accountability process in the 1980's and 1990's.

**History of Testing in North Carolina**

As the state of North Carolina's General Assembly pondered the idea of academic rigor and improving student graduation rates some 40 years ago, legislation was introduced that became the first formalized steps in North Carolina to assess student content mastery by measuring student proficiency through standardized testing. The General Assembly passed legislation in 1984 and 1989 that required the North Carolina Department of Public Instruction (NCDPI) to develop a core curriculum and evaluative assessments to gauge student progress. The first *End of Grade* (EOG) tests were rolled out in 1989. The tests were redesigned with educator input in 1992-1993 and then frequent test re-norming and curriculum updates occurred in subsequent years. In addition to the end-of-year assessments, the state created scale scores and the more familiar achievement levels for EOG tests and *End of Course* (EOC) tests. Four standards of achievement were ascribed to measure student success.

- **Level I** - Students performing at this level do not have sufficient mastery or knowledge and skills of the course or subject area to be successful in the next grade or content area**.**
- **Level -II** - Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject or content area and are minimally prepared to be successful at the next grade or content area.

- **Level III** - Students performing at this level consistently demonstrate mastery of grade level and content area subject matter and skills and are prepared for the next grade level or content area.
- **Level IV** - Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level and content level work. ("ABC Program Information," 2006)

In 1995, The North Carolina General Assembly enacted Senate Bill 16 directing the State Board of Education (SBE) to examine the administrative organization of the Department of Public Instruction (DPI) so that they might propose a plan for reorganization ("ABC Program Information," 2006). The resulting plan was entitled the *ABC's of Public Education*. The *ABC's* went into effect in North Carolina in 1996. With the creation of the North Carolina's *ABC's of Public Education*, the state required that schools annually report to the public two primary student accountability measures. Students scoring an achievement level of 3 or 4 on the EOG and EOC tests were grade-level or subject-level proficient. Unfortunately, by relying solely upon student test-proficiency percentages, North Carolina originally overlooked student growth in the respective grade levels or subject areas (Lachlan-Hache & Castro, 2015, p.3). From the outset, the *ABC's* of public education incentivized school performance (the percentage of students scoring levels 3 and 4 on EOG and EOC assessments) by granting financial rewards to schools for student test score excellence. However, the *ABC's* also provided consequences for not reaching state-prescribed goals. Schools designated as "Low-Performing Schools" were those that did not meet expected growth and had fewer than 50% of their students meeting an achievement level of 3 or above on EOG and EOC testing (ABC Program Information, 2006). Assistance teams from the state would be assigned to these schools to affect the change needed to grow student assessment scores. After a year of intervention, the team might recommend teacher and/or principal termination if appropriate steps were not taken to improve the school's achievement and growth results (NCDPI, 2001). In 2000 and 2001, in response to Congress's Individuals with Disabilities Education Act (IDEA) and the Bush administration's reauthorization of ESSA–the *No Child Left Behind Act* (NCLB), North Carolina began formally assessing all students including those identified as exceptional children, those participating in special education programs (NCDPI 2001). The bill gained support with both Republican and Democratic politicians as national concerns again arose around American schools' international competitiveness (Klein, 2015).

In addition to the light shone on student test score proficiency, NCLB also implored states to adopt growth models based on these same student scores. The legislation required that not only should student proficiency on end of year testing improve, but that subgroups of student scores should also improve. NCLB placed a special emphasis on advancing disadvantaged students within ethnic minority groups such as those served by the free and reduced lunch program, and subgroups of students with English as a second language (Klein, 2015). With the authorization of NCLB, the Federal government reinserted itself into public education, this time with greater influence and scope (Dee & Jacob, 2010; Klein, 2015). According to Dee and Jacob (2010):

> NCLB dramatically expanded the law's [ESSA's] scope by requiring that states introduce school-accountability systems that applied to all public schools and students in the state. NCLB require[d] annual testing of students in reading and mathematics in grades 3 through 8 (and at least once in grades 10 through 12) and that states rate

> schools, both as a whole and for key subgroups, with regard to whether they are making adequate yearly progress (AYP) toward their state's proficiency goals. (para. 2)

The Obama administration again reauthorized ESSA guidelines in 2009. In that year, the President and his administration linked a Department of Education grant program known as *Race to the Top* with an economic stimulus package, *The American Recovery and Reinvestment Act* (Klein, 2014). The administration used over six billion grant dollars to encourage states to embrace the administration's education policies. This included states allowing for the existence of charter schools, adopting a common curriculum with common end of year assessments, instituting college and career-ready standards, and reviewing teacher effectiveness using student test score analysis as one part of their evaluation (Klein, 2014). In order to reach the standards the DOE developed, many states began to tie the evaluation of teachers to the impact teachers were making on their students' learning. This impact was quantified using value-added-models or VAMs.

**Value Added Measures and EVAAS in the U.S., North Carolina**

Proficiency, according to Bandeira de Mello et al. (2009), measures a student's performance against an externally established benchmark that signifies mastery of a subject. In North Carolina, student test proficiency represents students scoring a level 3, 4, or 5 on an EOC or EOG. These summative examinations are created by trained teacher-test-writers from across the state familiar with the standards the tests are to measure. The questions are reviewed by other teachers and then added to a pool of questions imbedded as field test items on EOC or EOG tests. Each step in the process is important for maintaining instructional validity on the end of year suite of tests used in North Carolina's public schools (Pitre-Martin, 2018).

A student can show grade-level proficiency but not show growth. A model that would reward teachers with high levels of student proficiency is a proficiency model or "status model." Status or proficiency models report the status of a student by percentile, which illustrates a student's relative strength as compared to students across the state (Betebenner, 2009). Status models of performance do not take into account the growth a student can show on a given test even though they may not reach a level of grade-level proficiency. Student growth can be defined as the amount of academic progress a student makes in a school year (Cody, McFarland, Moore & Preston, 2010, p.2). Putting it another way, value-added-models share a common characteristic; they calculate a measure of change, or the value added to a student's learning over a period of time based on a student's performance on a standardized test (Sorenson, 2016).

Both achievement and growth measures are important in understanding a student's academic standing and in estimating how effective a teacher is in serving the individual needs of her students. Gauging teacher effectiveness based solely on achievement and growth measures assumes that all classroom student populations across the state, county, or even school are similarly constructed in terms of student demographics, general intelligence, and student academic readiness (Amrein-Beardsley & Collins 2012; Millman, 1997). Using only student test scores to gauge teacher effectiveness overlooks student growth that may not yet have risen to the level of grade-level competence (Lachlan-Hache & Castro, 2015, p.3). Carey (2004) summarizes the purpose for using student growth rather than proficiency as an indicator of school and teacher effectiveness:

> This means that teacher effectiveness measures aren't based on the overall performance of students; they're based on the improvement of students from the beginning of the school year to the end. A teacher who makes better than expected progress with previously low achieving students would be rated more effective than a teacher who made only moderate progress with previously high-achieving students, even if the high-achieving students still finished the year with better overall scores. (p. 5)

To address the high negative correlation between poverty and proficiency when assessing the effectiveness of schools and staffs, North Carolina found it necessary to measure student growth in addition to student test score proficiency (Jensen, 2009; Reardon, 2013).

Measuring teacher impact on student learning as measured by value-added- models did not begin in North Carolina. In Knoxville Tennessee, 1981, a University of Tennessee agricultural researcher, Dr. William Sanders, came across a newspaper article suggesting that there was no way for a school system to hold its teachers responsible for student test scores. But after reading the article, Sanders wrote a letter to the Tennessee Governor, Lamar Alexander, telling him that he could develop a statistical model that could measure teacher effectiveness by tracking student test score data (Aldrich, 2017). Up to this point, Sanders had devoted his research to agricultural output and wildlife issues. In the search for a more objective measure of teacher effectiveness than principal evaluations, he applied his statistical model for maximizing crop yields to the issue of teacher effectiveness as measured by the growth in student test scores. The Tennessee Value Added Assessment System (TVAAS) that Sanders developed, could assist education leaders in identifying the most effective and least effective teachers in what seemed a more objective way to measure teacher effectiveness than achievement levels alone (Sanders, 2003). According to Sanders and Horn (1998), the purpose of a VAM is to engender a climate for teacher self-reflection and improvement.

> If the purpose of educational evaluation is to improve the educational process, and if such improvement is characterized by improved academic growth of the student, then the inclusion of measures of the effectiveness of schools, school systems, and teachers in facilitating such growth is essential if the purpose is to be realized. Of these three, determining the effectiveness of individual teachers holds the most promise because, again and again, findings from TVAAS research show teacher effectiveness to be the most important factor in the academic growth of students. (Sanders & Horn, 1998, p.3)

By 1992, the Tennessee General Assembly adopted the Sanders Tennessee Value Added Assessment System (TVAAS) model as a core component of the *Education Improvement Act* (EIA) (Horn & Wilburn 2014, p. 70). The growth data this model produced, its creators claimed, showed how well teachers impacted their students' learning. The data was used to inform principals, teachers, and school district leaders until 2010 when TVAAS data began to be tied to teacher evaluations as well (Aldrich, 2017). According to Sanders, his model was created with the idea that teachers were the dominant factor in determining student growth, minimizing the importance of class size, student heterogeneity, student diversity and other factors that might influence student growth (Sanders, Wright, & Horn, 1998).

After Tennessee's VAM adoption, North Carolina was one of the next states to adopt a student growth component as part of their accountability measures. Average statewide growth was tabulated initially in 1993-94 in grades four through seven. This state growth statistic became the comparison point by which each school's growth was then measured (NCDPI, 2001). Grade three was added in 1997 when the third grade pretest was introduced and a pre-test/post-test growth model could be introduced. Grades eight and ten were finally added in 1997-1998.

The Department of Public Instruction felt growth was such an important component for evaluating teacher effectiveness that in 1998, the General Assembly passed a law stating that the State Board should develop performance standards to evaluate certified employees. These standards included improving student achievement. (N.C. Gen. Stat. §115C 335. 1998). The state determined that if a school's performance composite were less than 50 (percentage of children scoring a three or four on the end of grade reading and math tests) and the school's growth less than zero, then that the school would be designated as a "low performing" school (NCDPI, 2001). Trying to avoid this albatross, schools would aim to meet "expected" or "exemplary growth" (NCDPI, 2001). These growth distinctions were determined using a formula that examined how a school progressed over a baseline year. The scores of students in attendance for at least 91 days were averaged and compared against their aggregate average scale score from the prior year. The growth was calculated and reported at the school level (NCDPI, 2001).

Another key component in the development of the North Carolina testing program and growth model occurred in 2000-2001 when the State Board of Education weighted the *ABC's* growth composites to eliminate concern over small groups of students having a disparate impact on school growth (NCDPI, 2001). According to the NCDPI website chronicling the history of North Carolina's testing program:

> Rather than comparing different students from one year to the next, this plan – the School-Based Management and Accountability Program – holds schools accountable for the educational growth of the same students over time. By focusing on growth, the accountability model stimulates improvement for all schools. (NCDPI, 2001).

In 2005, North Carolina requested from the U.S. Department of Education (USED) the use of student growth as part of its reporting requirements for NCLB. Later the next year, North Carolina was approved for USED's *Growth Model Pilot*. For state accountability purposes, North Carolina's public school system used a trajectory model that calculated growth in grades three through seven using EOG tests for math and reading with the goal of all students in North Carolina growing sufficiently each year so that they would achieve on-grade-level status in both math and reading by eighth grade (WestEd, 2012). In 2009, NCDPI included growth measures as part of its application for the Federal *Race to the Top* grant monies that netted the state close to $400 million to use for education initiatives across the state (Cody et. al, 2010).

Most North Carolina teachers are now evaluated on the impact they make on their students' learning as measured by student growth on state assessments. The North Carolina public school system's most recent iteration of a growth model is the VAM known as the Educator Value-Added Assessment System (EVAAS) developed to measure a teacher's impact on the learning of the students she serves. EVAAS, the North Carolina version of Sanders' original TVAAS model, is now a product of the business statistics corporation known as SAS, based in Cary, North Carolina. Beginning in the 2011-2012 school year, EVAAS was adopted by the North Carolina State Board of Education. Starting in 2012, teacher growth indices were to be included in the North Carolina Educator Evaluation

System for both teachers and school administrators. This data became Standard VI on the teacher's evaluation in 2012-2013, Standard VIII for the administrator. A rolling EVAAS index average of the best two of three years for both teachers and administrators would determine their overall effectiveness.

EVAAS though is just one of several value-added models used to evaluate teacher effectiveness across the country. Hermann, Walsh, and Isenberg, describe a value added model or VAM as follows:

> VAM's predict individual student achievement based on the student's characteristics, including baseline achievement, and compare this prediction with the actual achievement of a teacher's students. The prediction is derived using data on other students in the state or district and represents what we would expect the student to achieve if taught by the average teacher. The difference between how a teacher's students actually performed and how they were predicted to perform represents the estimate of the teacher's value-added to student achievement. (2016, p.2)

According to the American Educational Research Association (AERA), VAM's can be divided into four categories.

1. Growth models, also called gain score–based or mean gain models, which simply aggregate difference scores derived from subtracting previous scores from current scores on tests;
2. Transition-based models, or categorical models, which compute aggregate changes in performance categories over a period of 2 or more years;
3. Student growth percentiles–based models, which answer the question "What is the percentile rank of a student's current test score compared to students with

similar previous test scores?" The individual teacher's VAM score is then calculated as the median or mean percentiles aggregated across her or his students;

4. Value-added measures–based models (VAM), which establish an expected current test score for students based on test scores from previous years, along with (possibly) other demographic characteristics of the student, classroom, and the school in attempting to account for the impact of factors beyond student achievement to isolate the teacher's impact (AERA, 2015).

Today, the EVAAS instrument is used not only in North Carolina but also in Pennsylvania, Houston, Ohio, and Tennessee to measure the impact of districts, schools, and teachers on the learning of their students. By 2014, a variety of VAM models were being used in 44 states (Collins & Amrein-Beardsley, 2014). According to a fact sheet from their own website, SAS proclaims of its VAM instrument:

> EVAAS is the most comprehensive reporting package of value-added metrics available in the educational market… Through the breadth of reporting and interactive nature of web delivery, districts and schools can set their own policy goals and assess their success. (SAS, 2017a, para. 2)

Using growth metrics to measure employee effectiveness may seem a better fit for the industrial or agricultural sector than it does in public education considering that the purpose of the Sanders' model originally was to predict agricultural outputs (Amrein-Beardsley, 2014). However, some prominent politicians on both the left and right have supported the idea of growth data being used to evaluate teacher effectiveness. In a 2009 speech, President Barack Obama made the case for linking educator effectiveness to student data.

> Success should be judged by results, and data is [sic] a powerful tool to determine results. We can't ignore facts. We can't ignore data. That's why any state that makes it unlawful to link student progress to teacher evaluations will have to change its ways if it wants to compete for a [Race to the Top] grant. (para. 1)

Despite Sanders' declaration that the primary use of EVAAS data was to be for teacher reflection and self-improvement, VAM's today are used for more than just teacher reflection in places where it has been adopted (Sanders & Horn, 1998). State legislatures and educators frequently employ EVAAS and other VAM data to make termination decisions, to compensate teachers, and to guide principal evaluations (Goldring, Grissom, Rubin, Neumerski, Cannata, Drake, & Schuermann, 2015). Even in the early years of VAMs, critics worried about how these growth measures would be used despite Sanders best intentions for the data his methodology would provide. When an assessment system tries to serve two purposes, both the accountability and improvement functions, the system is less than optimal for accomplishing either purpose (Goldring et. al., 2015; Millman, 1997).

**Value Added Model Critiques**

William Sanders was optimistic about his work. He felt that he had developed a regression model that could be used to evaluate teacher effectiveness and at the same time remove the subjectivity of traditional teacher evaluation instruments like teacher observations and student test scores.

> An integral part of TVAAS is a massive, longitudinally merged database linking students and student outcomes to the schools and systems in which they are enrolled and to the teachers to whom they are assigned as they transition from grade to grade. Research conducted utilizing data from the TVAAS database has shown that

race, socioeconomic level, class size, and classroom heterogeneity are poor predictors of student academic growth. Rather, the effectiveness of the teacher is the major determinant of student academic progress...For these reasons, a component linking teacher effectiveness to student outcomes is a necessary part of any effective educational evaluation system. (Sanders & Horn, 1998, p.1)

Despite political support for using VAMs like Sanders' TVAAS to measure student academic success and in turn teacher effectiveness, several academicians and researchers have found evidence to provide them pause in anointing TVAAS, North Carolina's EVAAS, or other VAMs as the panacea to educator evaluation problems (Amrein-Bearsley, Pivovarova, & Geiger, 2016; Collins, & Amrein-Beardsley, 2014). Researched and debated in education circles, the validity and reliability of value-added models that supply growth indexes for many public school teachers across the United States are being called into question over their use as instruments to measure teacher success. Not only have critics of VAMs expressed reservation about using student growth models for teacher evaluation, they have also been critical of the accuracy of these measures considering the myriad of student characteristics/variables that are outside the control of the classroom teacher. As early as 1973, Brophy argued against the use of achievement tests and student test score gains to assign teacher accountability. Sources of instability must be identified and controlled when using student growth as a measure of teacher effectiveness (Brophy, 1973, p. 251). One source of instability in VAM model measurements Harris (2010) identifies as systematic error. Specifically, Harris (2010) speaks of the "summer learning loss" that may affect some schools and students more than others, especially in communities with fewer libraries. More recently, the American Statistical Association reported that:

- VAMs are generally based on standardized test scores, and do not directly measure potential teacher contributions toward other student outcomes.
- VAMs typically measure correlation, not causation: Effects - positive or negative - attributed to a teacher may actually be caused by other factors that are not captured in the model.
- Under some conditions, VAM scores and rankings can change substantially when a different model or test is used, and a thorough analysis should be undertaken to evaluate the sensitivity of estimates to different models. (ASA, 2014).

Newton (2010) and her colleagues reviewed three different VAMs in their longitudinal study, but even in models where demographics are controlled, there were negative correlations across the board with teachers of English learners, free lunch recipients, and Hispanic students. Conversely, these same researchers found that there were positive correlations with Asian students and those students whose parents were highly educated (Newton, Darling-Hammond, Haertel, & Thomas, 2010). "However, it is also possible that a substantial share of what some would call a 'teacher effect' actually measures other factors that are correlated with student characteristics" (Newton et. al, 2010, p. 19). The authors challenge the fairness of using a VAM model that cannot take into effect student rate of learning influenced by prior knowledge, student attendance, time provided at home for homework, parent's tutoring, and even class sizes (Newton et al., 2010).

Notwithstanding the numerous critiques chronicled above concerning the use of educational VAM's emanating from academia and educational statisticians alike, states like Texas, New Hampshire, Pennsylvania, and North Carolina joined Tennessee in their search for a VAM tool to measure teacher effectiveness through student test data.

**Teacher Incentives and Merit-Based Pay**

Teacher incentive pay is not an idea unique to North Carolina policy makers. In the early 19th century, when the grade-level schooling model was developed, secondary teachers were paid more because it was thought that older students were more difficult to teach than younger ones (Guthrie, Springer, Rolle, & Houck, 2007). By the early 20th century, collective bargaining and the rise of labor unions returned teaching salaries to the more familiar step-based compensation model (Podgursky & Springer, 2007). In the 1920's and 1930's, merit pay briefly returned to favor when scientific management theory prevailed and administrators adopted business model evaluations of their staffs (Johnson & Papay, 2010). Cold war interest in merit pay emerged again in the 50's and 60's and sophisticated observation codes were instituted only to lose favor again in the 1970's (Johnson & Papay, 2010). Podursky and Springer (2007) found a brief resurgence in pay for performance models after a *Nation at Risk* was released in 1983, and today, merit pay plans are gaining momentum once again in part due to federal incentives to include teacher growth in state ESSA reauthorization plans.

In addition to merit pay initiatives in North Carolina since 2015, the following states and municipalities have instituted performance pay as well:

- Denver Public Schools' Professional Compensation System for Teachers (ProComp)
- Florida's Merit Award Program (MAP)
- Minnesota's Quality Compensation Program (Q-Comp),
- Texas' Governor's Educator Excellence Award Programs, and

- National programs like the Milken Family Foundation's Teacher Advancement Program (TAP) and the USDOE's Teacher Incentive Fund (TIF). (Pham, Nguyen, & Springer, 2017, p. 8)

In general, proponents of merit-based pay cite two primary reasons for their support of pay for performance.  First, merit pay encourages and motivates teachers to improve their efforts since these efforts will be rewarded (Springer & Taylor, 2016).  Consequently, employers do not have to designate how the outcome is to be achieved, but can allow for academic freedom in instructional methods, as long as student growth is occurring (Holmstrom & Milgrom, 1991).  Secondly, financial incentives may attract and retain higher performing teachers to the profession (Ballou & Podgursky, 1998).

In a cynical view of teacher merit pay, P.L. Thomas states:

> In education reform, the surveillance of students, and now the surveillance of teachers, is not covert, but in plain view in the form of tests and the resulting test data. Relying on tests and test scores for accountability allows that surveillance to be disembodied from those students and teachers (and thus to appear objective) and examined as if a reflection of merit. (Thomas, 2013 p. 89)

Despite Thomas's warning, North Carolina legislators have begun to favor teacher compensation reform plans that rely less on step plans that reward years of service and education attained, and more on teacher effectiveness as measured by student growth on end of year state-created exams (The Facts on Merit Pay for Teachers from Public Schools First NC, 2016; Eckert & Dabrowski, 2010).

In 2013 the North Carolina General Assembly directed the state's public school systems, local education associations (LEA's), to develop a process for rewarding the top

25% of teachers in the their LEA with a bonus of $5000 to be given over four years with the caveat that those that receive it agree to give up tenure rights (NC Sess. Laws 360, 2013; Thomas, 2013). The state did not set limits on how the LEA's would identify this top 25%. However, responding to *The Excellent Schools Act* of 2013, the ESSA legislation established,

- Plans to implement a system of merit pay for North Carolina teachers
- A commitment to a "robust evaluation instrument" and further assessments of teachers.
- A specific focus on student growth as a factor for teacher evaluation (*Public Schools First NC's* "The Facts on Merit Pay for Teachers," 2016).

However, the state public school superintendents and the North Carolina Association of Educators widely rejected the merits of this bill; it was repealed before it went into effect.

Eckert and Dabrowski in their 2010 research supported the idea of tying teacher compensation to student performance. In order to move forward on this pay per performance model, they proposed three questions as it related to teacher effectiveness and VAMs in general:

- What are the current limits of value-added measures?
- What combination of measures could be used to determine teaching effectiveness?
- How can we move away from divisive rhetoric toward necessary teacher compensation reform? (Eckert & Dabrowski, 2010, p. 88-89)

The proposed changes Eckert and Dabrowski documented in teacher compensation reflected the General Assembly's movement toward a merit-based pay system once again.

To understand the push for recent performance bonuses in North Carolina, one must understand the Read to Achieve (RTA) law (2012). Part of the greater *Excellent Public*

*Schools Act*, the RTA law was designed to ensure that all students are reading at or above grade level by the end of third grade (Wagner, 2014). In 2016, the NC General Assembly passed SL 2016-94 (Section 9.7) in an attempt to reward North Carolina's third grade teachers that most impacted student learning. The purpose of the compensation bonuses as described in the legislation was to study the effect of these incentives on teacher performance and retention (SL 2016-94).

The free market reform efforts to transform teacher compensation models that flowed from session law 2016-94 continued in the fall of 2016 when the North Carolina General Assembly allocated $5 million for teacher bonuses for both the 2016-2017 and 2017-2018 school years. Possibly in an attempt to attract strong teachers to third grade, these bonuses were allocated to the top 25% of third grade teachers in the state according to their EVAAS indexes. Another reading bonus was to be given to those teachers finishing in the top 25% of third grade teachers in their LEA (SL 2016-94, 2013). As a result, some third grade teachers in the state earned merit pay exceeding $7000 (SL 2016-94, 2013).

In Section 8.8 of SL 2016-94, the NC General Assembly allocated bonuses to additional public school teachers. SL 2016-94 allowed for teachers of Advanced Placement (AP) courses to earn $50 per student that scored a 3, 4, or 5 on the associated AP tests. Teachers of International Baccalaureate (IB) classes also received bonuses for fours and higher on IB examinations. Finally, in Section 8.9 of the same law, high school Career and Technical Education Teachers (CTE) could earn $25 or $50 for children who earned industry certifications or credentials. One stipulation for these high school teachers was that no AP, IB, or CTE teacher could receive more than $2000 in merit pay in a year.

The state also allocated in this same bill $17,242,627 for a "Merit Based Bonus for non- Educators." Local Education Agencies (LEAs) developed local plans on how this money could be used for classified staff such as custodians, clerical staff or teachers' assistants. They could also choose to develop criteria to distribute these bonus dollars for principals, assistant principals and central office personnel as the LEA saw fit (SL 2016-94.)

For any of these merit-pay initiatives to meet the burden of fairness to all teachers eligible to receive it, the VAM instrument identifying these effective teachers must yield accurate and valid results. One concern for teachers in using their VAM data to determine teacher effectiveness is the fact that the growth data the VAMs produce from summative student assessments is often not returned until the fall of the following year. In a survey completed in 2014, only 10% of North Carolina teachers responded that performance-based pay would incentive teachers to work more effectively, and almost nine in ten felt that performance pay would negatively affect teacher morale and teacher retention, ruining the collaborative nature of teachers in public education (Oakes & Robertson, 2014). In addition, there are many different subject areas, especially in North Carolina high schools, where growth measures are not calculated and for many subject areas and grade levels, compensation is not tied to their EVAAS growth (Goldring et. al., 2015.) These courses include many Career Technical Education classes, foreign languages, health and physical education, music courses, and the other arts. Teachers of these courses receive no EVAAS data. Most importantly, even for those teachers that receive a growth measure, what specificity do these growth indexes provide in informing a teacher as to what they are doing that positively impacted student learning? (Goldring et. al., 2015, p. 96).

**Student Growth Related to Student Characteristics and Non-Random Assignment**

Despite the many studies that critique the methodology used by VAMs, William Sanders' EVAAS model receives its greatest criticisms from researchers who find that the lack of random student assignment nullifies any validity to the growth results his model calculates (Amrein-Beardsley, 2014). Sanders has made the claim that his model adequately accounts for all the potential influences on student growth (thereby allowing the isolation of teacher direct effects) on learning, by employing the experimental design principle of "blocking," using each student's prior achievement as the only control or "proxy" for all such influences (Kupermintz, Shepard, & Linn, 2001, p. 6).

> [E]ach child can be thought of as a 'blocking factor' that enables the estimation of school system, school, and teacher effects free of the socio-economic confoundings that historically have rendered unfair any attempt to compare districts and schools based on the inappropriate comparison of group means. (Sanders, Saxton, & Horn, 1997, p. 138)

According to Kupermintz et al. (2001), due to the lack of student random assignment to teachers in most schools, blocking out "contaminating" variability is suspect, not allowing for the separating out of student achievement and teacher effectiveness (p. 7-8). Sanders would counter that his VAM model that compares students against their own testing history, nested within state results on end of year tests, makes it unnecessary to control for extraneous variables because the students' characteristics are already accounted for in their previous test scores (Sanders et. al., 2009). It is these "extraneous variables" that Sanders (2009) references that cause the most concern for researchers who reviewed EVAAS results.

One of the harshest critics of VAMs is Audrey Amrein-Beardsley. In her work, she identifies *5 Sticky Issues* that provide the basis for her VAM critiques. She questions the *reliability* and *validity* of VAMs. She finds VAM results to be *biased* and that the models lack *transparency*. Finally, she finds fault with VAMs in what she found to be a lack of *fairness* (Amrein-Beardsley, Pivovarova, & Geiger, 2016, p. 37). In so far as student characteristics may have an effect on teacher growth outside what Sanders' model can account for, Amrein Beardsley's sticky issues categories of *bias* and *fairness* may provide a framework for reviewing student characteristic effects on teacher growth results (Amrein-Beardsley et. al., 2016).

As Kupermintz et al. (2004) explained in their evaluation of TVAAS, the model was built to be used in controlled experiments that require students be randomly assigned to classrooms. Paulfler and Amrein-Beardsley found in their 2014 study, however, that random assignment of students to classrooms is not common practice within schools. Newton and her colleagues concluded in their 2010 study that if schooling systems are not set up to allow random assignment between students, teachers, and schools, then no statistical model, regardless of how sophisticated it is, can accurately measure teacher effectiveness. Braun (2005) found similarly that no statistical model, however complex, and no method of analysis can account for causal attribution if there is no randomization in the sample.

David Monk (1987) interviewed hundreds of principals across 17 east coast states to find out what they considered when they assigned students to classes. Most reported random assignment within student groupings. Principals would assign them "randomly" within achievement, behavioral, learning styles and student demographics groups (Dills & Mulholland, 2010; Monk, 1987). Others admitted that they would allow teachers who knew the kids best to make the assignments. It is also common for principals to use ethnicity, gender, student behavior, language proficiency and parental requests to inform principal assignment practices (Burns & Mason, 1995). Researchers have also found that principals will reward teachers by assigning higher achieving students, students with limited behavioral problems, fewer English as a Second Language (ESL) students, and students not burdened with poverty (Player, 2010).

In a 2012 study of VAM models that focuses on the validity of allowing teacher growth to be included in their evaluations, the authors indicate several problems related to student characteristics and how they are tied to teacher performance in terms of growth (Darling-Hammond, Amrein-Beardsley, & Rothstein, 2012). Differential summer learning loss particularly affects low-income children, and most end of year state tests, typically written on grade level, are incapable of measuring student learning below and above grade level with any level of accuracy (Darling-Hammond et. al, 2012; Harris, 2010). This would make accurate accounting of academically gifted student growth difficult to calculate. The authors continue by finding that when students aren't randomly assigned to classrooms, no VAM can fully adjust for the disproportionate numbers of impoverished students or those with English as a second language (ESL) that may be seated in one teacher's classroom (Darling-Hammond, Amrein-Beardsley, & Rothstein 2012; Collins & Amrein-Beardsley, 2014).

Collins and Amrein-Beardsley found that a preponderance of ESL students, students with disabilities, and even academically gifted students negatively bias teacher growth estimates, the latter because of the ceiling effect that even Sanders and Horn (EVAAS creators) acknowledged in their writing (Collins & Amrein-Beardsley, 2014; Sanders, Saxton, & Horn, 1997). Other authors have found a statistically significant correlation between peer ability groups. High achieving students, when grouped with other high achieving students, tend to grow at a higher rate compared to heterogeneous student groupings. This phenomena is known in research circles as the *peer effect* (Burke & Sass, 2013; Sacerdote, 2011). Other researchers found that ALL students groups benefited when peer grouped with high achieving students (Hanushek, Kain, Markman, & Rivkin, 2003). The authors of both studies found negative peer effects when classrooms were populated with low achieving students. These student groupings tended to retard the growth of low achieving students (Burke & Sass, 2013; Sacerdote, 2011). Sacerdote (2011) found statistically significant results in classrooms where same-race peers' starting achievement levels were similar.

"No statistical manipulation can assure fair comparisons of teachers working in very different schools, with very different students, under very different conditions" (Haertel, 2013, p. 24). Haertel seems to be speaking to every teacher in North Carolina who finds special challenges with the group of students they serve. However, with this reality in mind, there are competing opinions about how teachers should be compensated. According to the 2011-2012 Schools and Staffing Survey, 95% of public school districts in American have a single salary schedule for compensating teachers (Pham, Nguyen, & Springer, 2017). However, researchers have found that there is little correlation between teachers' years of service and student performance, leaving policy makers and school administrators to ponder whether compensating teachers based on years of service is the best use of local and state dollars, or whether a system of merit pay, incentive pay, performance pay, or pay for performance as it is sometimes called, is a better compensation model (Hanushek, 2003;

Podgursky & Springer, 2010; Springer, 2009). Merit pay offers an alternative teacher compensation system for a tax paying public that may have concerns that teacher compensation based solely on degrees earned or years of service weakens teachers' incentive to exert more effort into their work (Pham, Nguyen, & Springer, 2017).

Several studies have noted that merit pay is situated within principal-agent theory where theoretically, workers are motivated by monetary incentives, when individual efforts are not easily measured (Dixit, 2002; Heinrich & Marschke, 2010; Holmstrom & Milgrom, 1991). Agency theory assumes that people want to avoid risk or hard work and a supervisor must account for this by creating a compensation system that compels the worker to work while minimizing risk (Jensen, 1983). This theory is inadequate for studying merit pay in the current public education system because numerous studies have found teachers to be motivated by reasons other than personal economics (Besley & Ghatak, 2005; Cohen & Murnane, 1985; Horne, Foley & Flora, 2014; Kellough & Lu, 1993; Marsden & Richardson, 1994). The overarching assumption in agency theory is that agents, or teachers, need an economic reason to show up to work every day when researchers have found that teachers are also motivated by non-economic variables and intrinsically motivated as well (Horne et al, 2014).

**Conceptual Framework**

Critiques of VAM models and their use in high stakes decisions involving teacher dismissal and merit pay are not hard to find with the simplest of Google searches. Two studies, however, serve as a jumping off point for my research. Baker et al. (2010) outline the weaknesses of all VAMs including North Carolina's EVAAS metric. The authors summarize their findings this way:

> While those who evaluate teachers could take student test scores over time into account, they should be fully aware of their limitations, and such scores should be only one element among many considered in teacher profiles. Some states are now considering plans that would give as much as 50% of the weight in teacher evaluation and compensation decisions to scores on existing poor-quality tests of basic skills in math and reading…but the serious problems of attribution and nonrandom assignment of students, as well as the practical problems described above, would still argue for serious limits on the use of test scores for teacher evaluation (p. 20).

The researchers' focus on non-random student assignment to classrooms and their concern over teacher compensation tied to VAM scores will help shape my research.

The second study that will inform my methodology will be the Burke and Sass study from 2013 in which the authors reviewed student peer effects using linear and non-linear models to gage their influence on individual student achievement gains. Using the student demographic and socio-economic status variables described in Chapter One, I will design a model that discovers if teachers with different student clusters are truly receiving appropriate consideration when it comes to their opportunity to receive a merit-based bonus.

**Synthesis of Literature**

The stakes are high for North Carolina's elementary and middle school teachers with the onset of merit pay. With the opportunity in grades three through eight to earn up to $7000 in additional salary, teacher EVAAS growth measures are more important today than ever for teachers, if not for students. Even though the educator growth indexes are no longer tied directly to teacher evaluations, the North Carolina legislature is now compensating teachers who are able to grow their students most effectively, independent of the students' reaching grade level proficiency as measured by EOG math and reading assessments. The researchers who have tackled the VAM issues tied to teacher evaluation, teacher compensation, and teacher dismissal are predominantly critical of the high stakes use of EVAAS and other VAMs. The models' inability to control for non-random student assignment and peer effects on student performance remain unaccounted for in the model despite the "blocking" effects claimed by EVAAS model proponents that mitigate any of the concerns the critics cite. Due to the conflicting ideas on the VAM models' measurement validity, a quantitative analysis of teacher merit-based compensation in grades three through eight in relation to their EVAAS indices will be a study worth doing.

**Summary**

In Chapter Three, I will describe an exploratory model that utilizes a quantitative methodology to look at teacher impact on student learning (growth) in a rural western North Carolina County that have had students non-randomly assigned to their classrooms. In this study, I will gather the bonus data on those teachers in 2017-2018 who received the merit pay in 20 elementary and middle schools that serve students in grades three through eight. The sample size of students in math and reading will be over 5000 children who tested in both math and reading during the 2017-2018 academic year. I will review growth indexes for approximately 220 teachers that serve these students. Finally, I will develop a logistic stepwise regression model that will reveal if students characteristics prevalent in a teacher's classroom are having any influence on which teachers are receiving merit-based bonuses for most effectively impacting their students growth as measured by the EVAAS growth instrument. If the EVAAS instrument is sufficiently robust without the use of student characteristic variables to identify effective teachers, I would expect no student characteristics to appear as significant and impactful in predicting which teachers will receive merit-based bonuses.

## Chapter 3: Research Method and Design

### Introduction

There are teachers who work with the impoverished, schools serving mostly second language learners, school systems where minority populations prevail: Do these teachers and schools share the same opportunity to achieve expected growth or to exceed growth as schools that serve more advantaged populations? Alternatively, is it possible or even likely, that the teachers of students in poverty have a greater chance to achieve high growth indexes due to their students having more room to grow? With this growth conundrum as a primary driver of my research, I will develop in this chapter a methodology that will best answer the research question − is the likelihood of a North Carolina teacher receiving incentive pay impacted by the student characteristics prevalent in her classroom?

### Research Design and Paradigm

Creswell defines and explains the nuances of three different research designs in his writings. These research methods consist of qualitative, quantitative, and mixed (Creswell, 2005). In writing *The Practice of Social Research*, Babbie (2010) describes research using a quantitative methodology this way:

> Quantitative methods emphasize objective measurements and the statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques. Quantitative research focuses on gathering numerical data and generalizing it across groups of people or to explain a particular phenomenon. (p. 84)

Following Babbie and Creswell's quantitative research guidelines, the plan for my research is to gather pre-existing data on student characteristics, exclusively categorical in nature, to help explain why certain teacher receive the bonuses for reading and math instruction in grades three through eight as described in the research question. The goal is for this study to be effectively valid and sufficiently reliable because the method design is repeatable and the findings are generalizable in part or in whole to other districts across the state of North Carolina (Babbie, 2010; Creswell, 2005).

Embedded within the concept of a quantitative research framework is what Creswell refers to as a worldview and what other theorists have called research paradigms (Creswell, 2005; Guba & Lincoln, 2008). The general philosophical orientation, paradigm, or worldview that I bring to this research is one of post-positivism. This paradigm is an offshoot from the positivist philosophical framework derived from the scientific method more prominent among researchers in the early to mid-1900’s (Creswell, 2005). Positivism is the idea that there is a truth that can be found by quantifying human behavior through data and research (Phillips & Burbules, 2000). The post positivists, however, are not as dogmatic in demanding their data and subsequent research findings reveal an absolute truth. Nonetheless, the post positivists do seek to identify the causes that influence outcomes (Creswell, 2005; Phillips & Burbules, 2000). A post positivist would posit that although our knowledge is fallible we should not quit seeking the truth (Phillips & Burbules, 2000). Creswell (2005) posits that post positivism is reductionist. The intent of post positivist research then is to reduce ideas into a small, discrete data set to test, such as the variables that comprise hypotheses and research questions (Creswell, 2005). Finally, post positivism is deterministic in philosophy in that it assumes there are causes that determine outcomes (Creswell, 2005).

When choosing a quantitative research design, a researcher must select one that takes into consideration the population from which the data is drawn and the type of data that can be obtained before selecting a model that best explains a phenomena being examined (Creswell, 2005). As previously stated, the research methodology for this study will be quantitative. According to Creswell, a quantitative approach is one in which:

> …[T]he investigator primarily uses postpositivist claims for developing knowledge (i.e., cause and effect thinking, reduction to specific variables and hypotheses and questions, use of measurement and observation, and the test of theories), employs strategies of inquiry such as experiments and surveys, and collects data on predetermined instruments that yield statistical data. (2005, p.8)

Rather than surveys or experimental data, this quantitative study will use data that represents student characteristics in grade three through eight classrooms and the corresponding growth data of the teachers that served these students. The research will be descriptive rather than experimental as this data already exists and no treatments will be applied to the subjects of the study as in experimental research (Creswell, 2005). This research will also be exploratory. I will identify student characteristics that have the greatest impact on teachers' likelihood of receiving a merit-based bonus, or I may find that no variables influence a teacher's likelihood to receive the merit bonus. Through the post positivist lens however, I will conduct my research not to assume causality between the independent and dependent variables, but rather to find associations and relationships among the predictor and outcome variables. I will look for correlations between the dependent variables, teachers receiving the bonus or not for math and reading instruction, and the 13 independent variables that describe the student characteristics used in this study (Table 3.1).

When taking on a quantitative research study with thousands of pieces of data, the researcher must follow certain guidelines as outlined by Babbie (2010). My research methods will be documented alongside Babbie's guidelines for quantitative research.

- *The data is usually gathered using structured research instruments.* This research will guide the collection of existing student demographic and SES data from the North Carolina Secure Shell server via the PowerSchool student information system portal.
- *The results are based on larger sample sizes that are representative of the population.* This research will consider the characteristics of *all* students in grades three through eight in the selected county of study. However, this is not necessarily a representative sample of all similar-aged peers from across the state of North Carolina.
- *The research study can usually be replicated or repeated, given its high reliability*. This idea of repeatability will be discussed in the validity and reliability section of this chapter.
- *The researcher has a clearly defined research question to which objective answers are sought*. Is the likelihood of a North Carolina teacher receiving incentive pay impacted by the student characteristics prevalent in his/her classroom?
- *All aspects of the study are carefully designed before data is collected.* The design of this research is to use a logistic regression forward step-wise approach for the binary independent and dependent variables in the study.
- *Data are in the form of numbers and statistics, often arranged in tables, charts, figures, or other non-textual forms.* The data is categorical in nature, so it is not continuous like data that is analyzed through multiple and simple linear regression.

The SPSS tool will run the logistic regression and the findings will be reported with data tables.

- *The project can be used to generalize concepts more widely, predict future results, or investigate causal relationships.* This research will hopefully spur others in the field to replicate its processes to see if different regions of the state reveal similar findings. This might also be replicated in other states where merit-based bonuses are being utilized and incentive pay considered (Johnson & Papay, 2010, Player, 2010; Springer & Taylor, 2016; Thomas, 2013).
- *The researcher uses tools, such as questionnaires or computer software, to collect numerical data*. The data collected here will have been gathered from the North Carolina Department of Testing and Accountability's Secure Shell that links to the PowerSchool student database and information system (Babbie, 2010).

**Logistic Regression**

This study will employ a logistic regression model that will use nominal categorical binary variables that will be identified later. Categorical data can be nominal or ordinal. Ordinal data can be compared and ranked as in 1$^{st}$, 2$^{nd}$, and 3$^{rd}$. Nominal data is only differentiated by its name, like a group's ethnicity or a person's enrollment in a program (McColl, 2019).

In general, logistic regression seeks to:

- **Model** the probability of an event occurring depending on the values of the independent variables, which can be categorical or numerical
- **Estimate** the probability that an event occurs for a randomly selected observation versus the probability that the event does not occur

- **Predict** the effect of a series of variables on a binary response variable
- **Classify** observations by estimating the probability that an observation is in a particular category (such as approved or not approved). (Foltz, 2015)

I will have to reject other types of linear and non-linear regression models because the data I will be using is categorical and binary in nature. Therefore, the data will not have a normal distribution needed for most typical linear regressions (Foltz, 2015).

In the vein of post positivist thinking where no absolute truth can ever be found, I will develop a hypothesis and a null hypothesis, and then rather than fully except the hypothesis on its validity, I will seek to reject the null (Phillips & Burbules, 2000). *The **null hypothesis** for this study is that there is no impactful and statistically significant student characteristic variable, to a .05 significance level, that accounts for a teacher having a better chance than another teacher to earn a merit-based bonus in math or reading based on the composition of students in her classroom.*

**Design Rationale**

Creswell (2005) tells us that if a researcher finds a problem that leads him to identify factors that influence an outcome, then a quantitative approach is best (p.22). This particular research will look at the state's merit-based pay program for teachers of students in grade three through eight in the subjects of reading and math. The program was designed to encourage and reward teaching excellence and to encourage better personnel to enter into the teaching profession (*The Facts on Merit Pay for Teachers*, 2016).

Continuous data like test scores, IQ, height, weight, age, etc. may be better suited to a linear or multiple regression model. The data used for this study, however, will be categorical or binary in nature. The data will not be continuous. It will be nominal rather than ordinal since there will be no degrees in the data like a researcher may find with survey responses using Likert scales (Creswell, 2005, McColl, 2019). For example, one binary categorical variable used in this study is a child's involvement in the Exceptional Children's program. Either the child is served by the Exceptional Children's program or he is not. Another example is the English Language Learner variable. Either a child is served by the ELL program or she is not. There are no degrees to ethnicity although there are six categories of race that will all be coded with a zero or one. In this research, even poverty is a binary predictor variable as the child is either receiving free or reduced lunch or he isn't. Because the student characteristics will be either yes or no in relation to the variables, the study will incorporate nominal categorical predictive (independent) variables (McColl, 2019). The dependent variable, teachers receiving the bonus, will be binary as well. The teacher must be in the top 25% of teachers across the state in their given subject and grade to receive the bonus based on their EVAAS growth data.

In a forward stepwise approach, I will seek to accept or abandon some of the categorical variables as valid predictors of the dependent variable. Following the step-wise approach of adding and removing variables, I will find a model of best fit that illustrates which student characteristics, if any, are significant predictors of teachers receiving the merit-based bonus. Because of a binary dependent variable, plotting the predictor variables versus the response variable does not yield linear scatter plots. Logistic models do not use simple linear regressions as a result, but in order to develop predictive models, the data must be adjusted so as to mimic a linear regression. In order to regress y onto $x_1$, one must convert the dependent binary variable into a logit that looks like this: $Ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k$ where the logit function is the natural log of the odds that Y equals one of the categories (in this case, the probability is .25 that a teacher will receive the bonus), and *P* is defined as the probability that Y=1 (Logistic regression with Stata Chapter 1, 2019). The *X's* are the independent predictor variables (the student characteristics in this model), and the $\beta$ is the beta coefficient of the independent variables. In logistic regression, the $\beta$ indicates the amount of change expected in the log odds when there is a one-unit change in the predictor variable with all of the other variables in the model held constant (Logistic regression with Stata Chapter 1, 2019).

One of the concerns I have in my data sets is the multi-collinearity of independent variables. I foresee correlations between several of the variables and the Socio-Economic Status (*SES*) variable, represented in this study by students served by the free and reduced lunch program. The choice of a logistic regression model and a stepwise process help to mitigate this concern of multi-colinnearity as logistic models are designed to compensate for this statistical phenomona. Questions about the collinearity of variables will be discussed in Chapter Five under recommendations for future research.

**Role of the Researcher**

To protect student anonymity, I will interact with the data only after it has been masked. I will request through appropriate school district protocol that the rural school district's employee with access to the state's Secure Shell (where student testing data is stored) retrieve the data that will contain student test scores and student characteristics. The same employee will extract the teacher growth data that can be found in the EVAAS portal, match it against payroll information designating which teachers received the EVAAS-dependent merit-based bonus, and merge the two data files to connect teacher merit-pay data with student data. From the combined data set provided by the district, I will use forward step wise logistic regression in order to find any student characteristic variable(s) that shows a strong correlation to teachers that exceeded expected growth and thereby earned the math or reading bonus allocated to the top 25% of teachers in those grades and subjects in North Carolina's public schools.

**Ethical Issues**

The ethical issues in my research will revolve around the gathering of data and the anonymity provided using student demographic and student characteristic data. I will also look at data representing teachers who have received merit-based bonuses. I do not supervise any of the teachers whose data is being considered for this study. The school district will gather all of the data that I will analyze before I put my hands on it. The student names and identification numbers and teacher names will be removed so that I will be unaware of any of the individuals represented in the data set. Biasing the data should be minimal since I will not be interacting within the data collection environment as one might if observing a classroom or interviewing a subject (Creswell, 2005, p. 171). The student and teacher information will be collected from existing databases and will not be altered.

**Data Sources**

The data will come from several state testing databases. The North Carolina Department of Public Instruction (NCDPI) uses a Secure Shell to transfer and store all test scores and student demographic data. Secure Shell receives its student demographic and characteristic data from the student information system in North Carolina known as PowerSchool, from software associated with the Exceptional Children's program, and from data entered by teachers into the NCTest Admin portal used to provide accommodations to students for online testing.

## Data Collection

The data will be downloaded from the state server or Secure Shell described previously. NCDPI pulls this data from PowerSchool, the student data information system that houses attendance, grades, parent access portals, student demographic information, and all other relevant educational information for North Carolina public school students. The state compiles this student information into a file on Secure Shell called PCAUDIT. This is short for "Proficiency Audit" and it catalogs all EOG and EOC student scores as they progress from third through twelfth grade and beyond. In addition to student scores, the file attaches student demographic data to their test scores in grades three through eleven.

## Participants

The goal of quantitative research studies is to be able to make generalizations about larger groups based on the findings of a smaller population of study participants (Blackstone, 2012; Creswell, 2005). Different techniques for sampling the larger population are conducted in quantitative research. Probability sampling and representative sampling aim to gather participant samples that are representative of the larger group. Random selection techniques are also utilized to be sure that the group to be studied possesses proportional samplings of the larger group's characteristics so that valid and reliable generalizations might be drawn from study findings (Blackstone, 2012).

In this study however, I am fortunate to have access to almost 100% of the population of students in a rural western North Carolina county in grades three through eight that completed the math and reading EOG's in the 2017-2018 school year. In Chapter Five I will discuss limits to generalizability that manifest not because the participants are not correctly represented in the data, but because the populations of students in other LEA's carry with them student characteristics represented in different proportions.  All students testing in reading in grades three through five and all students testing in math in grades four through eight will be part of the study. This includes approximately 5000 students and about 8000 test scores from the *North Carolina End of Grade* assessments in the corresponding grades and subject areas. Students will be characterized with the following demographics and personal characteristic variables (Table 3.1).

Table 3.1

*Study Participants*

| Participant Group Categories | Description |
|---|---|
| Ethnicity | Asian, African Americans, Hispanic, Mixed Race, Pacific Islander, White |
| Socio-economic level | The data available for this characteristic is dependent on the child's participation in the free and reduced lunch program. |
| Sex | Male or Female |
| English as a second language program participant | Student participation in the English as a Second Language or EL program |
| Academically and intellectually gifted | Students identified for and participating in the AIG program. |
| Exceptional children's program participants | Students served by the special education program.  Those served in this instance are not those students being instructed on the *Extended Content Standards (*Extend 1 students*)*, but those that participate in the regular curriculum and are on track to receive a regular high school diploma.  These students may be served in speech, occupational or physical therapy, for having a reading disability, a math or writing disability, or being "other-health-impaired."  The "other health-impaired" special education label includes students with "attention deficit hyperactivity" disorder, and other health concerns that may affect their ability to access the curriculum. |
| Teachers receiving the bonus | In North Carolina, the top 25% of teachers in grades four through eight received a bonus in the 2017-2018 academic year as a result of their students growing academically in math.  In grades three through five, the top 25% of teachers impacting student learning as measured by EVAAS also earned a bonus. |

## Participant Selection

In order to access a robust data base, the study will examine student testing results from the 2017-2018 school year in EOG reading and math scores for students in grades three through eight. Any summer school testing (second test administrations) will not be considered for this study. Only student first-test-administration results from May and June of 2018 will be considered for this research since it is only the first test results that are used to tabulate teacher growth data. All students testing in this rural western North Carolina county will be included in the study. I will use data from 21 schools across six grade levels in two subject areas to expand my data set to over 5000 students and their subsequent characteristics.

## Institutional Review Board Procedures

Institutional Review Boards (IRB) are situated in research institutions to satisfy Federal requirements for ethical research but also, as summed up the Plymouth State University IRB web site:

- To promote the safety and well-being of human participants
- To ensure adherence to the ethical values and principals underlying research
- To ensure that only ethical and scientifically valid research is implemented
- To allay concerns by the general public about the responsible conduct of research. (2019)

The Appalachian State University Internal Review Board serves a valuable institutional and fundamental research function. The IRB serves to protect human subjects and manage their risk during their involvement in research conducted in conjunction with the university. As part of the review process, I had to complete online training modules to be granted permission by the University and the IRB to conduct my study. A central theme to the tutorials was that the researcher must at all times maintain student confidentiality. These courses would set the framework for how I could gather data, conduct the research, and report the findings. Since I would be working with existing data and would not be interacting with subjects, the primary concern of the IRB was that I would maintain student anonymity within my data collection and publication. In fact, to maintain a level of separation from the data collection process, a representative of the LEA collected the data, and then removed any identifying student markings such as student names and identifying numbers before I could receive it. The “Agreement to Use Student Data” letter (Appendix I) was submitted to the Appalachian State University IRB board by my dissertation chair to summarize my research and any risk to subjects involved therein.

The superintendent’s permission to conduct this study was also gained although the letter of permission will not be printed here to maintain the anonymity of the LEA where the data was gathered. My dissertation chair and the Appalachian State IRB reviewed and approved my research proposal within the parameters of access granted by the superintendent on behalf of the school board. Results of the study will be shared with this same superintendent and school board upon completion although again, individual student data will be destroyed upon study completion.

**Data Coding**

In alignment with the logistic regression procedures described above, the dependent variable, teachers having received the merit-based bonus or not, will be binary. Each of the independent variables will be binary as well. Those teachers having received the bonus will be coded with a “1” for the students that they served. Those who did not will be a “0”.

Table 3.2 outlines the dependent and independent variables, the type of variable each is, the variable symbol, and the way that the variable will be coded.

Table 3.2

*Dependent and Independent Variable Data Coding*

| Variable | Variable Type | Variable Symbol | Coding |
|---|---|---|---|
| Teachers receiving the reading bonus (Gr. 3-5) | Dependent | *RDGB_1* | 1 for teacher receiving bonus, 0 for not receiving |
| * Teachers not receiving the reading bonus (Gr. 3-5) | Dependent | *RDGB_0* | 1 for teachers not receiving the bonus, 0 for receiving |
| Teachers receiving the math bonus (Gr. 4-8) | Dependent | *MTHB_1* | 1 for teacher receiving bonus, 0 for not receiving |
| * Teachers not receiving the math bonus (Gr. 4-8) | Dependent | *MTHB_0* | 1 for teachers not receiving the bonus, 0 for receiving |
| Ethnicity | Independent | *Asian* | 1 if a student is Asian, 0 if they are not |
| | Independent | *Black* | 1 if a student is African American, 0 if they are not |
| | Independent | *Hispanic* | 1 if a student is Hispanic, 0 if they is not |
| | Independent | *Indian* | 1 if student is Native American, 0 if they is not |
| | Independent | *Mixed* | 1 if a student is mixed race, 0 if they is not |
| | Independent | *PacificIs* | 1 if a student is pacific islander, 0 if they are not |
| | Independent | *White* | 1 if a student is white, 0 if they are not |
| Sex | Independent | *Male* | 1 if a student is male, 0 if they are not |
| | Independent | *Female* | 1 if a student is female, 0 if they are not |
| Academically and Intellectually Gifted | Independent | *AIG* | 1 if a student is in the AIG program, 0 if they are not |
| ** Students with Disabilities | Independent | *SWD* | 1 if they are served by the SWD program, 0 if they are not. |
| ***Socio-Economic Status | Independent | *SES* | 1 if a student receives free or reduced lunch, 0 if they do not |
| English Language Learner program (new comers and long term EL's) | Independent | *ELL* | 1 if a student is served by the ELL program, 0 if they do not |

** The reason for this reverse coding is due to the negative coefficient statistic found when running the logistic regression. See findings in Chapter 4.*
*** For this study, students that are served by the AIG program will NOT be included in the SWD category as they are in some state level reporting that combines both into an Exceptional Children's category.*
**** The SES variable is a proxy for poverty in this study*

### Student Population Groups

Conducting research that yields valid and reliable results requires sufficient sampling of the overall population to be studied. Research populations consist of a group of individuals comprised of the same characteristics (Creswell, 2005, p. 595). In this case, students in grades three through eight that took the math and reading EOG's in the 2017-2018 school year makeup the research population. Problems can arise in interpreting findings from the research and in generalizing those findings to the larger group if the researcher is unable to draw a representative sample from the entire population (Creswell, 2005, p. 597). In this study, I will use close to 100% of the students in the population as my sample. I will be looking at the characteristics of the students that tested in grades three through eight in the identified school system during the 2017-2018 school year. Only a small number of students (less than 10 in both the math and reading populations) were discarded due to their not having an assigned reading or math teacher when data was compiled for this study. The frequency table below (Table 3.3) illustrates the number of students that fall into each category (predictor variables) and the percentage of the student population they account for in this exploratory logistic regression study.

Table 3.3

*Frequency Table for Math and Reading Bonus, Student Population*

| Student Characteristics | Math Students Grades 4-8 | % of Math Students | Reading Students Grades 3-5 | % of Reading Students |
|---|---|---|---|---|
| *Total* | *4490* | | *2925* | |
| Asian | 277 | 6.2% | 178 | 6.1% |
| African-American | 205 | 4.6% | 131 | 4.5% |
| Hispanic | 779 | 17.3% | 515 | 17.6% |
| Native American | 3 | 0.07% | 2 | 0.07% |
| Mixed Race | 286 | 6.4% | 189 | 6.5% |
| Pacific Islander | 7 | 0.14% | 2 | 0.07% |
| White | 2933 | 65.3% | 1908 | 65.2% |
| Male | 2318 | 51.6% | 1551 | 53.0% |
| Female | 2172 | 48.4% | 1374 | 47.0% |
| Academically and Intellectually Gifted (math and/or reading) | 864 | 19.2% | 469 | 16.0% |
| Students with Disabilities | 728 | 16.2% | 526 | 18.0% |
| Served by Free and Reduced Lunch Program | 2748 | 61.2% | 1884 | 64.4% |
| English Language Learner | 311 | 6.9% | 308 | 10.5% |

**Data Analysis**

A forward stepwise logistic regression model was employed to find a model of best fit in determining whether to accept the hypothesis or not. The ***Null hypothesis (*** $H_0$) *for this study is that there is no impactful and statistically significant student characteristic variable, to a .05 significance level, that accounts for a teacher having a better chance than another teacher to earn a merit-based bonus in math or reading based on the composition of students in her classroom.* To understand the soundness of the logistical regression model and its statistical findings, Peng, Lee and Ingersoll (2002) recommend analyzing the following:

a) The overall model evaluation

b) The statistical tests of individual predictors

c) A goodness-of-fit statistics and

d) The validations of predicted probabilities (p. 5)

For overall model evaluation, three inferential statistical tests: the likelihood ratio, score, and Wald tests will serve as indicators as to whether the model is an improvement over the null hypothesis (Peng et al. 2002).

The SPSS program will produce several statistics to help me evaluate the viability of the null hypothesis. To test the significance of the individual predictors (testing their regression coefficients, $B$), using the $Wald\ \mathcal{X}^2$ statistic will indicate whether they are significant at the $p < .05$ level. We will look at the degrees of freedom statistic (*df)* and the odds ratio that SPSS reports as *Exp(B)* (Field, n.d.). This log ratio will provide information to the researcher about what correlations exists between students with certain characteristics, and the likelihood of teachers receiving the bonus in math or reading that work with these student characteristic groups. Said another way, since logistic measures are the natural log of odds ratios, they can be transformed back into probabilities to examine if high probabilities are associated with events and low probabilities with non-events (Peng et al. 2002).

When SPSS generates its tables, it will be important to read the sign of the coefficient (*B*). If the coefficient is negative, it is indicating that there may be an inverse relationship between the predictor variables and the dependents. In the case of the independent variables having a negative *B* coefficient, I will recode the dependent variable with a "1" for not receiving the bonus and a "0" for a teacher receiving the bonus. This should flip the sign of the dependent variable and change the *Exp(B)* statistic to something that is positive and from which information can be extrapolated. This new log ratio may indicate that certain student characteristics have the effect of decreasing the likelihood of a teacher receiving the math or reading bonus.

**Trustworthiness**

Being objective is an essential aspect of competent inquiry, and for this reason researchers must examine methods and conclusions for bias. For example, standards of validity and reliability are important in quantitative research (Creswell, 2005).

I will not be utilizing a previous researcher's instrument to gather data like I might if I were doing surveys or interviews. I will be creating a logistic regression best-fit model to analyze student characteristics that might influence a teacher's receiving the merit-based bonus. The reliability of my study will rest on its ability to be repeatable, and the findings generalizable. Limited research exists in North Carolina on subject-specific merit-bonuses tied to individual teacher growth in large part because the teacher bonus model has only been in place since the 2016-2017 school year. As a result, there may be some interest in repeating this research in other regions of the state where student demographics differ. A researcher may want to conduct a study to discover if the state model is the fairest way to identify teachers that should receive the performance bonuses. Other researchers may critique my findings and find that my model left out one or more important characteristics to which I may or may not have had access. In short, the trustworthiness of my study will hinge on its repeatability and the generalizability of its results.

## Chapter 4: Results of the Research

### Introduction

The North Carolina General Assembly has been adjusting the teacher salary schedule since the passing of the *Excellent Public Schools Act of 2013*. The state legislature has ratified statutes that have significantly altered the step approach to teacher pay, a system commonly used among school systems nationally that appropriates teachers yearly step raises for each year employed or for degrees earned, regardless of the teachers' effectiveness. The North Carolina teacher salary schedule implemented in July of 2018 has provided larger percentage raises for teachers in their first five years compared to teachers in years 15-24 who see no graduated pay, significantly altering the year by year raises evident in most public school teacher pay scales (Appendix D - North Carolina Public School Salary Schedules, 2018).

In an attempt to link teacher pay to teacher effectiveness, legislators hoped to incentivize teaching excellence by providing merit-based bonuses that would reward excellent teaching, subsidizing those teachers that were the most effective at impacting their students' learning (Helms, 2017). The 2018 salary scale (Appendix D) is a modified version of graduated teacher pay, but to improve their take home wages, teachers in grades three through eight can best supplement their incomes by being in the top 25% of teachers locally and in the state as measured by their EVAAS reading and math growth indices. Through an exploratory quantitative methodology, I will examine through logistical regression whether the characteristics of students in grades three through eight in a rural western North Carolina school district positively or negatively affect a teacher's likelihood of receiving the state bonus for teachers of reading and math in these specific grades.

**Results**

In order to conduct the research, the SPSS regression tool was used to regress the different student characteristics represented by the predictor variables, onto the dependent variables, the teachers either receiving the bonus for reading and math growth or not receiving the bonus. For the study I used a logistic regression model. Logistic regression is the preferred regression model for handling binary dependent variables (Peng, Lee, & Ingersoll, 2002; Wuensch, 2014). For the sake of consistency, both the dependent variables and the independent variables were constructed as binary categorical variables. As an example, a student who receives free and reduced lunch was coded as a "1," a student who pays for their lunch with a "0." A student who is served by the English Second Language program as a "1," a student who is not as a "0." This binary coding is true of each of the student characteristic and demographic data as defined in Chapter Three.

A forward stepwise methodology was utilized for this logistic regression. When selecting variables for any regression work, the goal of variable selection is parsimony (NCSS Statistical Software, p. 311-1). The goal according to NCSS is to "achieve a balance between simplicity (as few regressors as possible) and fit (as many regressors as needed)" (p. 311-1). With simplicity and fit as its aim, the forward stepwise approach starts with no predictor variables in the model and then adds variables as long as their *p-value* is less than the predetermined alpha (*a*), in this example, .05. In other words, those variables selected have less than a 5% chance of being related to the dependent variable merely because of chance (Creswell, 2005, p. 598). In determining a logistic regression model of best fit, I will be attempting to reject the null hypothesis outlined in Chapter Two. *The* ***null hypothesis*** *($H_0$) for this study is that there is no impactful and statistically significant student* *characteristic variable, to a .05 significance level, that accounts for a teacher having a better chance than another teacher to earn a merit-based bonus in math or reading based on the composition of students in her classroom.*

With a forward stepwise approach, the SPSS program continues to add variables as long as the variable has a *p-value* < .05, starting with the variable that has the greatest R-squared value and working its way through the predictor variables based on these two parameters (NCSS, p. 311.1).

The first logistic regression reflects the following conditions (Table 4.1.a-d):

1. The dependent variable, *RDGB_1* (Reading teachers having received the bonus or not), is coded as a "1" for a teacher receiving the bonus, "0" for a teacher that did not. Only 25% of reading teachers in the state received the merit-based bonus for having the top EVAAS scores in grades three through five.
2. Student data in grades six through eight was eliminated from this study since teachers serving kids in these grades were not eligible to receive the reading bonus by state statute. Only students in grades three through five with an assigned reading teacher were used for this reason.
3. Students taught on the "Extended Content" standards (took the Extend 1 end of year state assessment) were also excluded since these end-of-year assessments do not generate a growth statistic for the teacher. These students with special learning needs account for 1% of the student population and have special learning needs.

One can see from the "Variables in the Equation" (Table 4.1.a) that the forward stepwise regression was run four times. The "Variables in the Equation" illustrates which variable are significant at a *p-value* < .05 (Appendix E).

Table 4.1.a-d
*SPSS Binary Logistic Regression Results – Grades 3-5 Reading: Dependent Variable Coded as 1: (RDGB _1)*

Table 4.1.a
*Block 1: Method = Forward Stepwise (Conditional)*
*Omnibus Tests of Model Coefficients*

| | | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Step 4 | Step | 4.215 | 1 | .040 |
| | Block | 70.228 | 4 | .000 |
| | Model | 70.228 | 4 | .000 |

Table 4.1.b
*Classification Table*

| | | Observed | RDGB_1<br>.00 | Predicted<br>1.00 | Percentage<br>Correct |
|---|---|---|---|---|---|
| Step 4 | RDGB_1 | .00 | 756 | 822 | 47.9 |
| | RDGB_0 | 1.00 | 462 | 885 | 65.7 |
| | Overall Percentage | | | | 56.1 |

*a. The cut value is .500*

Table 4.1.c
*Variables in the Equation*

| | Ind. Var. | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 4 | Asian | .723 | .172 | 17.751 | 1 | .000 | ***2.062*** |
| | Mixed | .351 | .170 | 4.265 | 1 | .039 | ***1.420*** |
| | White | .714 | .095 | 56.327 | 1 | .000 | ***2.042*** |
| | SWD | -.301 | .099 | 9.245 | 1 | .002 | .740 |
| | Constant | -.643 | .085 | 57.765 | 1 | .000 | .526 |

a. Based on conditional parameter estimates

Table 4.1.d
*Variables Not in the Equation*

| Step 4 | Variables | Score | df | Sig. |
|---|---|---|---|---|
| | Black | .077 | 1 | .782 |
| | Hispanic | .021 | 1 | .884 |
| | Indian | .215 | 1 | .643 |
| | PacificIs | .215 | 1 | .643 |
| | F | .642 | 1 | .423 |
| | M | .642 | 1 | .423 |
| | AIG | .199 | 1 | .656 |
| | SES | .540 | 1 | .463 |
| | ELL | .088 | 1 | .766 |

*a. Residual Chi-Squares are not computed because of redundancies.*

The final best fit model produces the *White* variable statistics as follows, (*B* = .714, *S.E*. = .095, *Wald* = 56.33, *df* = 1, *Sig*. = .000, *Exp(B)* = 2.04) where *B* is the variable coefficient, *S.E.* is the standard error of measurement of the coefficient, the *Wald* statistic lets the researcher know if the *B* coefficient for that predictor is significantly different from zero (Field, n.d.). The *df* statistic indicates degrees of freedom, *Sig.* is the *p-value* or significance level and the *Exp(B)* is the log ratio which is important in assessing the level of correlation between the predictive and dependent variables. White students were significant at a *p-value* equal to .000. The *Exp(B)* statistic is the most telling of the *White* variable statistics. This statistic is interpreted to mean that white students are 2.04 times more likely than non-white students to be sitting in a classroom of a teacher that received the reading bonus.

The Asian student variable reveals a similar statistic regarding a teacher's likelihood of receiving the bonus. This variable yields the following statistics (*B* = .723 *S.E.* = .172, *Wald* = 22.76, *df* = 1, *Sig.* = .000, *Exp(B)* = 2.473). The *Exp(B)* statistic's log ratio effect is somewhat mitigated by the larger *S.E* statistic and smaller *Wald* even though a *p-value* of .000 makes the *Asian* variable a significant predictor. The lower *Wald* statistic and the larger *S.E.* associated with the Asian independent variable stems from the fact that only 178 Asian students are present in the study compared to over 1900 white students warning the researcher to interpret the *Asian* variable results more cautiously (Table 3.3). But even with the *Asian S.E.* being larger than is the *White* variable's, the *Asian* predictor variable is significant and points to a 2.06 times greater odds that an Asian student is sitting in a classroom with a reading teacher that received the bonus.

Another demographic category flagged by the SPSS model for its positive correlation in the reading regression was the mixed race student variable (*Mixed*). The statistics for mixed race students are (*B* = .351, *S.E.* = .170, *Wald* = 4.3, *df* = 1, *Sig*. = .039, *Exp(B)* = 1.42). Similar to the findings of Asian students, the margin of error (*S.E.*) is about twice as large as that for white students in large part due to the population size of mixed race students, 189. It is significant at the .039 level and the log ratio tells us that mixed race students are 1.4 times as likely to be sitting in the classroom of a teacher that received the reading bonus in grades three through five.

Finally, the SPSS forward stepwise regression applied the *SWD* or students with disabilities predictor variable into the model. The SPSS *SWD* statistic reads as follows: (*B* = -3.01, *S.E.* = .099, *Wald* = 9.22, *df* = 1, *Sig.* = .002, *Exp(B)* = .740) The negative *B* coefficient makes the *SWD* variable difficult if not impossible to interpret. It is significant at the .002 level but I will re-code the dependent variable in order to produce a statistic that is applicable and interpretable to the study. This is the last significant variable to include in the forward stepwise approach. In the Model Summary, (Appendix E) the "-2 Log Likelihood" is reduced with each addition of the identified predictor variables. The smaller the "-2 Log Likelihood" the better the model predicts the dependent variable. We can see in Step 4 of the "Variables not in the Equation" that there are no more significant variables to use in the regression since all are above the alpha significance level we set at .05 (Appendix E).

Table 4.2.a-d illustrates the logistic regression model for math teachers in grades four through eight with the following conditions.

1. The dependent variable for Model 4.2.a-d is *MTHB_1*. Once again for Table 4.2, the dependent variable is coded as a "1" for teachers receiving the merit-based bonus and "0" if the teacher did not.
2. There is no third grade math pretest as there is with reading (the Beginning of Grade 3 Test or BOG3). Without a third grade math pretest, there is no student data by which to generate an EVAAS growth index for teachers of third grade math, and therefore, they are not eligible to receive a merit-based bonus for their work.
3. Those students on the extended content standards taking the Extend 1 end of year exam in math are also not included in this data set as EVAAS does not calculate growth for the teachers of these students.

Table 4.2.a-d
*SPSS Binary Logistic Regression – Grades 4-8 Math Dependent Variable Bonus Coded as 1: (MTHB _1)*

Table 4.2.a
*Block 1: Method = Forward Stepwise (Conditional)*
*Omnibus Tests of Model Coefficients*

| | | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Step 5 | Step | 3.867 | 1 | .049 |
| | Block | 41.783 | 5 | .000 |
| | Model | 41.783 | 5 | .000 |

Table 4.2.b
*Classification Table*

| | | Observed | MTHB_1 .00 | Predicted 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 5 | MTHB_1 | .00 | 2670 | 68 | 97.5 |
| | | 1.00 | 1684 | 68 | 3.9 |
| | Overall Percentage | | | | 61.0 |

*a. The cut value is .500*

Table 4.2.c
*Variables in the Equation*

| | Ind. Var. | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 5[e] | Black | .462 | .146 | 10.079 | 1 | .001 | ***1.588*** |
| | Hispanic | .188 | .082 | 5.255 | 1 | .022 | ***1.207*** |
| | Mixed | .248 | .125 | 3.908 | 1 | .048 | ***1.281*** |
| | AIG | -.231 | .081 | 8.036 | 1 | .005 | .794 |
| | SWD | -.382 | .088 | 18.927 | 1 | .000 | .683 |
| | Constant | -.414 | .044 | 89.377 | 1 | .000 | .661 |

a. Variable(s) entered on step 1: SWD.
b. Variable(s) entered on step 2: AIG.
c. Variable(s) entered on step 3: Black.
d. Variable(s) entered on step 4: Hispanic.
e. Variable(s) entered on step 5: Mixed.

Table 4.2.d
*Variables Not in the Equation*

| Step 5 | Variables | Score | df | Sig. |
|---|---|---|---|---|
| | Asian | .723 | 1 | .395 |
| | Indian | 1.050 | 1 | .305 |
| | White | .673 | 1 | .412 |
| | PacificIs | .323 | 1 | .570 |
| | F | .004 | 1 | .950 |
| | M | .004 | 1 | .950 |
| | SES | 1.481 | 1 | .224 |
| | ELL | .069 | 1 | .793 |
| | Asian | .723 | 1 | .395 |

*a. Residual Chi-Squares are not computed because of redundancies.*

In the best fit math model (Table 4.2), several student characteristics manifest as significant. Students that self-identify as racially mixed (*Mixed*), African American (*Black*) and Hispanic (*Hispanic*) are significant demographic variables as are students with learning disabilities (*SWD*) and those identified as gifted (*AIG*). Each is significant at a *p-value* < .05. In step 1 of the forward stepwise approach, *SWD* is the first variable added to the constant. The *SWD* statistics, after all significant predictor variables are added, are (*B* = -.382, *S.E.* = .088, *Wald* = 18.93, *df* = 1, *Sig.* = .000 and *Exp(B)* = .683). As with the *SWD* variable in the *RDGB_1* regression (Table 4.1) a negative *B* coefficient makes this statistic unusable. However, because of its significance level, I will recode the dependent variable and rerun the forward stepwise logistic regression to see if there is a *Exp(B)* relationship between *SWD* and a teacher's likelihood of receiving a math bonus. Another explanatory non-demographic statistic that proves significant is students that have been identified for the Academically and Intellectually Gifted *(AIG)* program in the chosen school system. The statistics for *AIG* are (*B* = -.231, *S.E.* = .081, *Wald* = 8.04, *df* = 1, *Sig.* = .005 and *Exp(B)* = .794). There are 864 students in this student population that yields a relatively small *S.E.* statistic. However, like with the *SWD* predictor variable, the negative coefficient makes this an uninterpretable statistic. However, I will review this once again in Table 4.4 when the dependent variable coding is flipped.

Reviewing once again the frequency chart (Table 3.3), there are 286 students of an ethnically mixed race background (*Mixed*) used in the math teacher bonus logistic regression model. The *Mixed* variable was a significant predictor of teachers receiving the math bonus (*B* = .248, *S.E.* = .125, *Wald* = 3.91, *df* = 1, *Sig.* = .048 and *Exp(B)* = 1.28). The *Mixed* predictor variable has an *Exp(B)* predictive odds measurement of 1.28. This indicates that a mixed race child is 1.28 times more likely to have a teacher that received the math bonus than the average child. The African American variable (*Black*) statistics are (*B* = .462, *S.E.* = .146, *Wald* = 10.08, *df* = 1, *Sig.* = .001 and *Exp(B)* = 1.588). A black student is 1.6 times as likely to be sitting in a classroom where the teacher received a math bonus at the end of the year. An Hispanic child (*Hispanic*) reveals these statistics (*B* = .188, *S.E.* = .082, *Wald* = 5.23, *df* = 1, *Sig.* = .022 and *Exp(B)* = 1.21). We can read this to mean that an Hispanic child is 1.2 times as likely to be in a teacher's class that received the math bonus.

**Results Continued: Reversing the Dependent Variable Coding**

As a result of the first two forward stepwise models yielding statistics with negative coefficients, and log ratios, *Exp(B)* less than one, I altered the coding of the dependent variables by coding teachers that received the bonus with "0" and those who did not with a "1". This provides me with some descriptor variables yielding more meaningfully interpretive statistics that can inform my research question. The same conditions outlined in the first two models hold true in the reading Table 4.3 and math Table 4.4 with the only change being the inverse coding of the dependent variables.

Table 4.3.a-d
*SPSS Binary Logistic Regression – Grades 3-5 Reading: Dependent Variable Bonus Coded as 0: (RDGB _0)*

Table 4.3.a
*Block 1: Method = Forward Stepwise (Conditional)*
*Omnibus Tests of Model Coefficients*

| | | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Step 5 | Step | 3.867 | 1 | .049 |
| | Block | 41.783 | 5 | .000 |
| | Model | 41.783 | 5 | .000 |

Table 4.3.b
*Classification Table*

| | | Observed | RDGB_1 .00 | Predicted 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 4 | RDGB_0 | .00 | 885 | 462 | 65.7 |
| | | 1.00 | 822 | 756 | 47.9 |
| | Overall Percentage | | | | 56.1 |

*a. The cut value is .500*

Table 4.3.c
*Variables in the Equation*

| | Ind. Var. | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 4 | Asian | -.723 | .172 | 17.751 | 1 | .000 | .485 |
| | Mixed | -.351 | .170 | 4.265 | 1 | .039 | .704 |
| | White | -.714 | .095 | 56.327 | 1 | .000 | .490 |
| | SWD | .301 | .099 | 9.245 | 1 | .002 | 1.351 |
| | Constant | .643 | .085 | 57.765 | 1 | .000 | 1.902 |

a. Variable(s) entered on step 1: White.
b. Variable(s) entered on step 2: Asian.
c. Variable(s) entered on step 3: SWD.
d. Variable(s) entered on step 4: Mixed.

Table 4.3.d
*Variables Not in the Equation*

| Step 4 | Variables | Score | df | Sig. |
|---|---|---|---|---|
| | Black | .077 | 1 | .782 |
| | Hispanic | .021 | 1 | .884 |
| | Indian | .215 | 1 | .643 |
| | PacificIs | .215 | 1 | .643 |
| | F | .642 | 1 | .423 |
| | M | .642 | 1 | .423 |
| | AIG | .199 | 1 | .656 |
| | SES | .540 | 1 | .463 |
| | ELL | .088 | 1 | .766 |

a. *Residual Chi-Squares are not computed because of redundancies.*

As with Table 4.1, the Chi-square statistic (Appendix F) illustrates that by Step 4, 70.23% of the variation in the dependent variable is explained by the best-fit model. The sign of the *B* coefficients have flipped with this model where Asian and White students are negative, but this time, the *SWD* variable reveals a positive coefficient. The *Exp(B)* reflects an odds ratio of 1.35 for students with disabilities (*SWD*). Deciphering the odds ratio illustrates a 1.35 times likelihood that a student with a disability is sitting in the reading class of a teacher that ***did not*** receive the reading bonus. There are 526 *SWD* (Table 3.3) students that took the EOG reading test across the county in grade three through five in 2017-18 yielding a standard error of .099.

Table 4.4.a-d
*SPSS Binary Logistic Regression – Grades 4-8 Math: Dependent Variable Bonus Coded as 0: (MTHB _0)*

Table 4.4.a
*Block 1: Method = Forward Stepwise (Conditional)*
*Omnibus Tests of Model Coefficients*

| | | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Step 5 | Step | 3.867 | 1 | .049 |
| | Block | 41.783 | 5 | .000 |
| | Model | 41.783 | 5 | .000 |

Table 4.4.b
*Classification Table*

| | | Observed | MTHB_1 .00 | Predicted 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 5 | MTHB_0 | .00 | 2670 | 68 | 97.5 |
| | | 1.00 | 1684 | 68 | 3.9 |
| | Overall Percentage | | | | 61.0 |

*a. The cut value is .500*

Table 4.4.c
*Variables in the Equation*

| | Ind. Var. | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 5[e] | Black | -.462 | .146 | 10.079 | 1 | .001 | .630 |
| | Hispanic | -.188 | .082 | 5.255 | 1 | .022 | .829 |
| | Mixed | -.248 | .125 | 3.908 | 1 | .048 | .781 |
| | AIG | .231 | .081 | 8.036 | 1 | .005 | 1.260 |
| | SWD | .382 | .088 | 18.927 | 1 | .000 | 1.465 |
| | Constant | .414 | .044 | 89.377 | 1 | .000 | 1.513 |

a. Variable(s) entered on step 1: SWD.
b. Variable(s) entered on step 2: AIG.
c. Variable(s) entered on step 3: Black.
d. Variable(s) entered on step 4: Hispanic.
e. Variable(s) entered on step 5: Mixed.

Table 4.4.d
*Variables Not in the Equation*

| Step 5 | Variables | Score | df | Sig. |
|---|---|---|---|---|
| | Asian | .723 | 1 | .395 |
| | Indian | 1.050 | 1 | .305 |
| | White | .673 | 1 | .412 |
| | PacificIs | .323 | 1 | .570 |
| | F | .004 | 1 | .950 |
| | M | .004 | 1 | .950 |
| | SES | 1.481 | 1 | .224 |
| | ELL | .069 | 1 | .793 |
| | Asian | .723 | 1 | .395 |

*a. Residual Chi-Squares are not computed because of redundancies.*

To address the negative *B* coefficient in the Table 4.2 math regression, the dependent variable *MTHB_0* is recoded as a "0" for a teacher receiving the math bonus, and a "1" for a teacher that ***did not*** receive the bonus in grades four through eight. This reverse coding should yield a positive correlation between students that are labeled *AIG* or *SWD* and teachers not receiving the math bonus. Again, we see identical *Chi-square, p-value*, and *S.E.* statistics as we did with the math regression in Table 4.2. However, the sign of the *B* coefficients has flipped and the *Exp(B)* statistic is yielding an interpretable statistic for both variables. The *SWD* log ratio (*Exp(B)*) of 1.47 means that a student with a disability is 1.5 times as likely to be sitting in a math teacher's class that ***will not*** receive a bonus for her math instruction. A positive log ratio of 1.26 exists with the *Exp(B)* statistics for a gifted student *(AIG).* Here again we see a positive relationship between the AIG log ratio and a teacher ***not*** receiving the math bonus.

# Chapter 5: Conclusions

## Introduction

The postpositivist seeks to assess causes that influence outcomes as she tries to reduce ideas into testable hypothesis in the search to find truth and explain the world in which she lives (Creswell, 2005). In this postpositivist vein, Phillips and Barbules (2000) see the role of the researcher as one where she advances the relationship among variables by developing questions or hypothesis. As a postpositivist researcher trying to uncover truths, I will identify relationships and correlations in this chapter's analysis section between the dependent variable, teachers receiving merit-based bonuses in math and reading, and the independent predictor variables made up of student characteristic groups in this exploratory study.

Next, I will review the literature connections that eventually narrowed my focus of study to my research question. In the literature review, I identified gaps where a study of merit-based pay in North Carolina's public schools was warranted. Through this research process and data gathering, I have identified several limitations with the EVAAS instrument and the merit-based pay system based on EVAAS teacher growth indices. The research by Baker et al. (2010) on non-random student assignment and Burke and Sass' (2014) study which identified peer effects and their influence on student learning outcomes both served as conceptual frameworks for this study. These studies provide critiques of a VAM's ability to accurately identify effective teachers for high stakes decision-making including merit-based bonuses. I will use the lens provided by these frameworks to review the findings in my study. The implications of this research in the landscape of educational policy and merit-based pay proposals may warrant future research in this subject. These ideas will be discussed in greater detail throughout the chapter.

**Analysis-Literature Links**

Quantitative research is an inquiry-based approach useful for describing trends and explaining the relationship among variables found in the literature (Creswell, 2005). To conduct this inquiry the investigator specifies narrow questions, and analyzes numbers from the instruments using statistics (Mertler & Charles, 2005). From the results of these analyses, the researcher interprets the data using prior predictions and research studies. The final report, presented in a standard format, displays a researcher's objectivity and a lack of bias (Creswell, 2005, p. 597). Following Creswell's ideas on objectivity (2005), I have compiled and formatted the student characteristic data that serve as the focus of my study.

William Sanders, the founder of the TVAAS and EVAAS statistical models, has claimed throughout his VAM publications that a student's personal testing history provides its own controls when trying to isolate teacher impact on student learning that Sanders referred to as "blocking" (Sanders & Wright, 2009). Furthermore, Sanders and Horn (1998) made the claim that race, socioeconomic level, class size, and classroom heterogeneity are poor predictors of student academic growth (p. 1). It was Sanders' own research that informed the creation of my null hypothesis that there is no impactful and statistically significant student characteristic variable, to a .05 significance level, that accounts for a teacher having a better chance than another teacher to earn a merit-based bonus in math or reading based on the composition of students in her classroom. However, my research findings do not support the assumption of the null hypothesis. As an impartial researcher, I am ready to reject the null hypothesis as my findings reveal that both reading and math growth, as measured by EVAAS, are correlated to different student characteristics present in a teacher's classroom (Table 5.1). Is the likelihood of a North Carolina teacher receiving incentive pay impacted by the student characteristics prevalent in his/her classroom? The research question that drove my work can be answered in the affirmative. There are some student characteristics that are positively correlated and others that are negatively correlated with a teacher's likelihood to receive the reading and math merit-based bonuses in North Carolina.

Table 5.1

*Result Summaries for Significant and Impactful Variables - Math and Reading*

| Independent Variables | Subject | Correlation | Odds Ratio [Exp(B)] | Standard Error |
|---|---|---|---|---|
| Asian (Asian) | Reading | Positive | ***2.1*** | .172 |
| Mixed Race (Mixed) | Reading | Positive | ***1.4*** | .170 |
| White (White) | Reading | Positive | ***2.0*** | .095 |
| Students with Disabilities (SWD) | Reading | Negative | ***1.4*** | .099 |
| African-American (Black) | Math | Positive | ***1.6*** | .146 |
| Hispanic (Hispanic) | Math | Positive | ***1.2*** | .082 |
| Mixed Race (Mixed) | Math | Positive | ***1.3*** | .125 |
| Academically/Intellectually Gifted (AIG) | Math | Negative | ***1.5*** | .081 |
| Students with Disabilities (SWD) | Math | Negative | ***1.3*** | .088 |

Starting with the grade three through five reading teachers eligible to receive the state bonus for being in the top 25% of teachers in the state as measured by the EVAAS growth index, four predictor variables emerged as significant at a $p$-value < .05 (Table 5.1). These were the demographic variables of students being white, mixed-race or Asian, and the student characteristic of being served by the exceptional children's program (*SWD*).

Based on my findings, a reading teacher in grade three through five is more likely to receive the merit-based bonus if she has more mixed-race, white and Asian students in her classroom irrespective of her teaching talents. Newton et al. (2010) found in their research that VAM models like EVAAS showed positive correlations between Asian students and teachers whose students have come from educated parents. My research supports their findings on Asian students and their correlation with teacher growth. I do not have any data on parent education by which to confirm or challenge the Newton et al. study (2010). I do not have any specific data from the LEA where the study was conducted that would tie mixed-race or white students to a parent education variable in order to make a claim as to the reason these two variables are linked to teacher success.

The summer learning loss phenomenon was chronicled and studied by Darling-Hammond et al. (2012) as a reason that poorer students fall behind their peers academically, and why growth for these students manifests at a lower rate. However, under the EVAAS model, students are measured against their own testing histories year to year and teachers assume responsibility for these students' "growth" even though these students' learning may have regressed over the summer before they stepped foot in that teacher's classroom. "Summer learning loss" is associated with those students living in households of poverty (*SES*). In the county where I conducted my research, there is substantial overlap in the *SES* and *SWD* variables. Students that are served in the SWD program are often served by the free and reduced lunch program as well. This may explain in part why there is a negative association between teachers receiving the bonus and the students with learning disabilities sitting in their classrooms. Darling-Hammond and her colleagues (2012) also point to the fact that the end-of-year assessments are written on grade level although much of the instruction for *SWD* students happens below grade level. Summer learning loss and below grade level instruction may both be accounting for the negative correlation that exists between students with disabilities and the ability of their teachers to earn the state-allocated reading bonus.

With affluence comes a family's ability to provide tutoring for their child outside the school. This logic might manifest with the existence of a significant and negative correlation between the *SES* variable that serves as a proxy for student poverty, and the likelihood of a teacher earning the bonus. However, in the regression models for math and reading, I did not find this to be true (Newton et al., 2010). With this line of thinking, I would have also expected the *AIG* subgroup to be a significant and important variable in a teacher's likelihood of receiving the bonus as well, but this was not the case in either the reading or math regression models. On the contrary, with the math regression models, the *AIG* predictor variable proved to be negatively correlated with a teacher's ability to receive the bonus. This may point less to the *peer effect* espoused by Hanushek et al. (2003), and more to the *ceiling effect* chronicled in Collins and Amrein-Beardsley (2014) and Sanders, Saxton, and Horn's research studies (1997).

Baker et al. (2010) would argue that certain teachers are given preferential treatment in the students assigned to their classrooms when principals allot these teachers high performing students. Most teachers enjoy teaching gifted and talented students and may even petition their principal to teach classroom clusters of students with the gifted and talented label. The classroom discussions where gifted students are present, the gifted students' ability to grasp concepts more quickly, and the higher-level student interactions that happen in classrooms where these students reside are rewarding classroom dynamics for most teachers. On the other hand, teaching students with disabilities often proves difficult for many teachers. Implementing differentiated lesson planning and providing creative instructional strategies are challenges for teachers serving students with learning disabilities.

From the student's perspective, frustration can emerge as special education (*SWD*) students struggle to internalize complex learning concepts.

When Baker et al. (2010) referenced preferential student-teacher assignments, they were referring to teachers being assigned more high achievers in their classrooms and fewer students that struggled with their learning or struggled with the English language. My findings in the math regression do not support the conceptual framework findings of Baker et al. (2010) when it comes to gifted students since their assignment to a classroom actually hindered a teacher's likelihood of receiving the math bonus. The researchers' ideas on the *SWD* subgroups, however, were upheld as their assignment to a teacher's classroom worked against a teacher's quest to receive the state-appointed math bonus. On the other hand, if a student is being compared to their own testing history in EVAAS as Sanders et al. (1997) have laid out, then could the argument reasonably be made that these students have more room for growth then a gifted student already performing at the 97$^{th}$ percentile? As a reminder, this finding does not indicate that these *SWD* students are not performing on grade level (although most are not), only that their teachers are impacting their learning at a lesser percentage than their peers.

Of course, the quantitative researcher must be cautious in ascribing causal connections between variables. However, students with disabilities in my study are 1.5 times more likely to be sitting in a math teacher's class who ***did not*** receive the math bonus. This relationship exists despite all the learning supports these *SWD* students receive in this public school setting.

- They frequently work in a small group instructional setting.
- They often receive after school remediation.

- They have an individual education plan (IEP) that focuses their learning on prescribed and personal learning targets.
- They receive push in and/or pull out services from an exceptional children's teacher for a prescribed number of minutes each week in addition to the instruction they receive from their regular education teacher.
- The school must also comply with testing accommodations such as additional time and multiple test sessions for in class and end of year state assessments including the "test read aloud" option with the math assessment for students who have a reading or math disability.

In the county where this research was conducted, teachers are to provide 45 minutes of instruction in the core-reading program to all students. This instruction is delivered in heterogeneously grouped classrooms where students of all academic levels are working together. Is it reasonable to assume that positive peer effects would manifest as these *SWD* students work alongside and with gifted learners who model high level thinking, fluency, and articulation (Hanushek, Kain, Markman, & Rivkin, 2003)? With the employment of the SWD teaching strategies chronicled here we would expect to see growth with these at-risk students groups. However, even with these supports, it was not enough to overcome the learning issues they bring with them each day to the reading teacher's classroom. The findings of this study parallel those of Baker et al. (2010) concerning students with disabilities. My research reveals that the *SWD* subgroup has a negative correlation to their teachers' likelihood of receiving the North Carolina merit-based bonuses in reading. Students with the SWD label are 1.4 times as likely to be sitting in a teacher's classroom that ***did not*** receive the reading bonus.

The binary logistic regression yielded five significant and impactful variables when predicting the likelihood of a math teacher in grades four through eight receiving a bonus based on the student characteristic makeup of her class. For math, these significant and important variables were the ethnic variables *Mixed, Black,* and *Hispanic,* and the student characteristics *AIG* and *SWD*. Although less common than reading disabilities, some of the students that make up the SWD subgroup have a mathematics disability. However, students with math and reading disabilities on their IEP's or students who have an English Learner education plan due to their scoring less than a 5.0 on the WIDA ACCESS language proficiency test are often provided the testing accommodation known as read-aloud. This accommodation allows the computer, for online testing, or a teacher if designated in the personal educational plan, to read the EOG test questions aloud to the students. In theory, this accommodation permits student comprehension of the math word problems without struggling with the reading decoding challenges these questions can present on EOG math tests, especially to students with learning disabilities.

The SPSS math statistics show that Hispanic students are 1.2 times as likely to be in a teacher's classroom that received the math bonus. Although not all Hispanic students are served by the ELL program, about half of them are served in grades three through eight. The read-aloud accommodation may account in part for the positive correlation between Hispanic students and teachers receiving the math bonus. Of course, we might expect to see the ELL predictor variable to reveal a significant relationship with the dependent variable if this were the case, but the ELL variable is not significant in either study. One would reason this same logic would be reflected in the *SWD* predictor variable and it positively impacting a teacher's ability to receive the math bonus. However, actually the opposite is true. Students with disabilities, even with the read aloud math test accommodation, were actually 1.5 times more likely to be sitting in a classroom where the teachers ***did not*** receive the math bonus.

Carnoy and Garcia (2017) found that the overarching determinant of student learning as measured by test score proficiency was not race but socio-economic class. My study reveals in the math logistic regressions that historically poor subgroups, *Black, Hispanic*, and *Mixed* race students have from a 1.3 to 1.6 times greater likelihood of sitting in a classroom where a teacher received the bonus but the SES variable does not manifest as significant in the math or reading regression models. What might account for this phenomenon that seems to contradict Carnoy and Garcia (2017)? Only 286 students of 4490 fourth through eighth graders identified as mixed race, 779 as Hispanic, and 205 as African-American (Table 3.3). With larger standard errors (Table 4.4) these predictor variable results should be viewed with some level of caution in assigning any type of correlation let alone causation.

One independent variable appears to work counterintuitive to Burke and Sass (2013) in their *peer effect* research. The math logistic regression model (Table 4.4) reveals that *AIG* students have a negative correlation with their teacher's ability to receive the math bonus. They are 1.3 times as likely to be sitting in a math teacher's classroom that ***did not*** receive the bonus. It is safe to assume that the *ceiling effect* may be handicapping math teachers of gifted students from achieving high growth numbers in the county I studied (Collins & Amrein-Beardsley, 2014; Sanders, Saxton, & Horn, 1997).

Revisiting William Sanders' VAM ideas (Sanders et al,1997), he believed that it would not be necessary to include SES, demographic, or other student characteristic variables in his model because they were already baked into the students' individual testing histories. Sanders believed his model could get at the direct *teacher effects* on the students' learning by employing the experimental design principle of "blocking," which in essence used each student's prior achievement as the only proxy for future test results (Kupermintz, Shepard, & Linn, 2001, p. 6). The logistic regression work I completed in a western North Carolina rural county would challenge this blocking theory approach to the EVAAS growth model. If the blocking phenomenon were to hold true, I would have been able to accept the null hypothesis since no student characteristic variable would have showed a significance of $p < .05$. However, many predictor variables proved to be significant and impactful. As a result of this finding, I feel comfortable in saying that a student's own testing history is insufficient to accurately predict student growth, and subsequently, those student characteristics are influential in a teacher's likelihood of receiving the math and reading state-appointed bonuses. Consequently, rewarding North Carolina teachers based on the growth index calculated by EVAAS may be unduly influenced by the student characteristics present in her classroom.

**Addressing the Gaps**

While conducting my literature review for this study, I focused on three areas. The first was the viability of using the VAM tool EVAAS for a purpose for which it was not designed – assigning teacher merit pay. Sanders designed EVAAS to measure teacher impact on student learning regardless of the student's starting point since the student is being measured against his own testing history (2003). He hoped that educators would use the reporting mechanism EVAAS provides teachers to reflect on their teaching and to improve on their craft. Sanders hoped that teachers would dive into EVAAS' diagnostic reports that illustrate how teachers have served the different academic tertiles of students that cross through their classroom doors (Table 5.2) (Sanders & Horn, 2009). However, as several researchers have lamented, EVAAS has instead been utilized for high stakes decisions linked to salary supplements and teacher dismissal (Amrein-Bearsley, Pivovarova, & Geiger, 2016; Amrein-Beardsley, & Collins, 2012; Eckert, & Dabrowski, 2010).

I also focused a part of my research on student assignment to classrooms. A supposition underlying all VAM models is a random assignment of students to classrooms. Research shows us that this is not what principals do however when making student classroom assignments (Baker et al. 2010; Dixit, 2002; Heinrich & Marschke, 2010; Holmstrom & Milgrom, 1991).

Finally, I looked at teacher merit-based pay and its origins in public education as a means to motivate teachers, retain talented ones, and attract the best and brightest to the practice (Ballou & Podgursky, 1998; Springer & Taylor, 2016). However, I was unable to find research specific to North Carolina on the performance-based pay that was instituted in the state with the 2016-2017 school year and the fairness by which these bonuses were applied. Although merit-based pay is available also for career-technical education teachers and teachers of Advanced Placement high school courses, I focused my research on the math and reading bonuses allocated by the NC General Assembly Session Law in grades three through eight. The purpose for focusing on grades three through eight lies in the fact that ***all*** students in these grades take the EOG reading and math tests which provides a perfect cross section of the population, unlike Advanced Placement courses where teacher growth is not calculated and where the courses are catered to the most accomplished students. CTE courses are self-selected and do not offer the same sampling of students characteristics to conduct a meaningful study. Since studies were scarce or non-existent on the topic of merit-based pay, specifically a quantitative review of the way student characteristics affect a teacher's likelihood of receiving the merit-based bonus in North Carolina, I identified this gap as something worthy of study. As a result, I developed a quantitative methodology that would assess in part the fairness of the merit-based pay, and moreover a methodology that would answer my research question. Is the likelihood of a North Carolina teacher receiving incentive pay impacted by the student characteristics prevalent in his/her classroom?

**Limitations**

There are several limitations that I have identified while cleaning the data, running the regressions, and analyzing the results. One is in the way that the student data system assigns students to reading teachers. Reading teachers are matched backed to their students in the state testing software (WinScan) for test score reporting purposes. The student data portal, PowerSchool, identifies reading teachers as the Grade_X_Reading (X being the grade of the students) reading teacher. This PowerSchool designation of reading teacher is typically the "teacher-directed reading" teacher. The teacher-directed reading block is the period of instructional time when a teacher teaches a class of heterogeneously grouped students working from a core-reading program or basal written on grade level. However, in the county where the research was conducted, students also participate in ability-grouped guided-reading groups. They listen and interact daily with their teacher during teacher read-alouds, and they have self-sustained silent reading that often involves teacher-student conferencing. Students may be served by their homeroom teacher the entire day. This is common in third grade. However, the students might be served by more than one of the teachers in the grade level, especially in fourth and fifth grades. Yet the data that was collected and provided to me has one reading teacher assigned to one student. Although one reading teacher is primarily tied to a student in the state bonus structure, several have an effect on his reading growth. This is a prime example of the *spillage* effect introduced in Chapter One (Corcoran, 2010; Koedel 2009, Yuan, 2015).

The EVAAS tool does permit several teachers to claim instructional responsibility for students even though they are not assigned as the student's teacher in the teacher-directed reading block. However, the data available to me only shows teachers assigned to students through this teacher-directed match. In other words, the students in grades three through eight are often served by multiple reading teachers, but only one teacher is assigned to one student in this study.

In the four regression models, there are some variables that are manifesting as significant and impactful but have a small population as can be seen in Table 3.3 – *Frequency Table for Math and Reading Bonus, Student Population*. The *Mixed, Black, and Hispanic* variables in math and the *Asian* variable with the reading regression are the obvious examples. How should these small population independent predictor variables be interpreted? Is there an insufficient *N* that causes a large *S.E.* that should produce caution in interpreting the results?

Another limitation to the research findings relates to the results being generalizable and repeatable? The methods ***are*** repeatable, but I don't know that the results would be. If the findings yielded different significant and important variables in a similar study in Carteret County or in Charlotte, does that mean that my methodology is not sound or valid? I do not think that it would invalidate the methodology since there are so many cultural factors, professional development opportunities, population differences, and teaching talent variances between LEA's across the state. However, if the study is repeated elsewhere, and the impactful and significant variables differ, this too adds credence to this study. The EVAAS results would again prove a bias toward a particular student characteristic set. Therefore, the bonus system tied to Sanders' VAM that assumes student testing histories are sufficient to construct teacher growth measures is flawed for its use in assigning teacher merit pay (Darling-Hammond, Amrein-Beardsley, & Rothstein, 2012; Sanders, 1998).

**Revisiting the Conceptual Framework**

The conceptual framework for this study was based on two studies. The first was the work of Baker et al. (2010) and their concerns about VAM model efficacy outside the bounds of random student assignment. I also considered the work of Burke and Sass (2013) who studied positive peer effects that occur for lower and average achieving students when they are grouped with gifted students, and the negative peer effects that occur when lower achieving students are homogenously grouped during the bulk of their instructional time. Homogenous student grouping is a more common practice in the middle schools represented in this study where students are often tracked with similar-ability peers for most of their instructional day. It is less prevalent in elementary schools, although it happens there also. I took the non-random assignment of students as a given in this rural county. Students are often clustered in classrooms by their ELL classifications, by their learning disabilities, and by parent requests for teachers. Since non-random student assignment is more common than not, my research would reveal if Sanders ideas on "blocking" would be sufficient to not disadvantage any teacher with the student characteristics prevalent in her classroom.

Baker et al. (2010) expressed concerns over merit-based pay structures where students were not randomly assigned to class. A compliment to this theory is the principal-agent theory (Dixit, 2002; Heinrich & Marschke, 2010; Holmstrom & Milgrom, 1991) that suggests principals are purposely not assigning students randomly as they try to maximize student outputs, assigning student to teachers based on teaching style, student acumen, or parent requests. The researchers' concerns appear justified as I found five independent variables in math and four in reading that were significant and positively or negatively correlated with teachers' likelihood of receiving the reading and math bonus. The ceiling effect appears to have manifested with AIG students in the math regression where teachers of these bright students were handicapped from achieving sufficiently high EVAAS indexes necessary to qualify for the merit-based bonus (Collins & Amrein-Beardsley, 2014; Sanders & Horn, 1998). The positive peer effects documented by Hanushek et al. (2003) for students lucky enough to be with the academically gifted student cluster and the teachers who reap the growth rewards are not factored into the EVAAS model, nor are the negative peer effects when clusters of low achieving students populate a classroom (Burke & Sass, 2013; Sacerdote, 2011). In neither study did the *SES* variable show up as significant that might prove some peer effect was prevalent in improving the growth of students in poverty. In the study, AIG students only surfaced as significant and important in the math model, and here they were negatively correlated with teacher growth. It would be impossible to decipher from this study if the presence of gifted students helped to raise the performance of other students in their classrooms assuming that heterogeneity of student abilities is common in these classrooms. Therefore, it may be that the ceiling effect manifested in these math classrooms. The AIG student subgroup did not appear to advantage the teachers that served them in the way of merit bonuses.

Although it would be difficult to separate out the peer effects and the non-random student assignments in this quantitative study, a future mixed methods study where principal surveys and interviews were coupled with the quantitative analysis may further inform the research question related to teacher merit-based bonuses and how they are related to principal classroom assignment practices.

**Implications of the Research**

A study, if its results are valid and reliable, should have generalizable findings. At the district level, this means that the findings might be applied to current county policy and procedures. Today, there are more procedures than there are policies related to student-teacher assignment in the county where this study was conducted. The LEA has contracted with Harvard and Columbia Universities over the last three years to provide training to a sample of its teachers and administrators in the art of differentiated instruction, focusing primarily on how to plan and provide instruction for their neediest learners. There has also been a push in elementary school for small group guided-math instruction that target learning needs for students' in their specific areas of deficit. Several of the elementary schools participated in an after school grant that allowed for 60 minutes of guided instruction provided by quality math and reading teachers to serve at-risk students through a McKinney-Vento federal grant and additional grants written through NCDPI.

In this LEA there are stipulations for state and locally provided teacher training that must be successfully completed if those teachers are to serve AIG students in grades three through eight. These teachers would have completed local and state training modules that qualify them for the job of teaching gifted students. This AIG certification hurdle often handcuffs principals who may otherwise randomly assign these AIG students to grade level classrooms. ELL and SWD students are frequently grouped for the purpose of inclusion co-teaching in most schools across the LEA. The inclusion model works best when an ELL or SWD teacher comes into the regular education classroom and works alongside the classroom teacher in a co-teaching model, that ideally supplements student learning opportunities. It is hard to conduct push-in services for these ELL and SWD students if they are spread throughout the grade level. The inclusion teachers cannot be in two classrooms at one time to serve all their students' IEP-specific learning needs. This is yet another reason why non-random student clustering often occurs in these classrooms.

**Implications at the LEA Level**

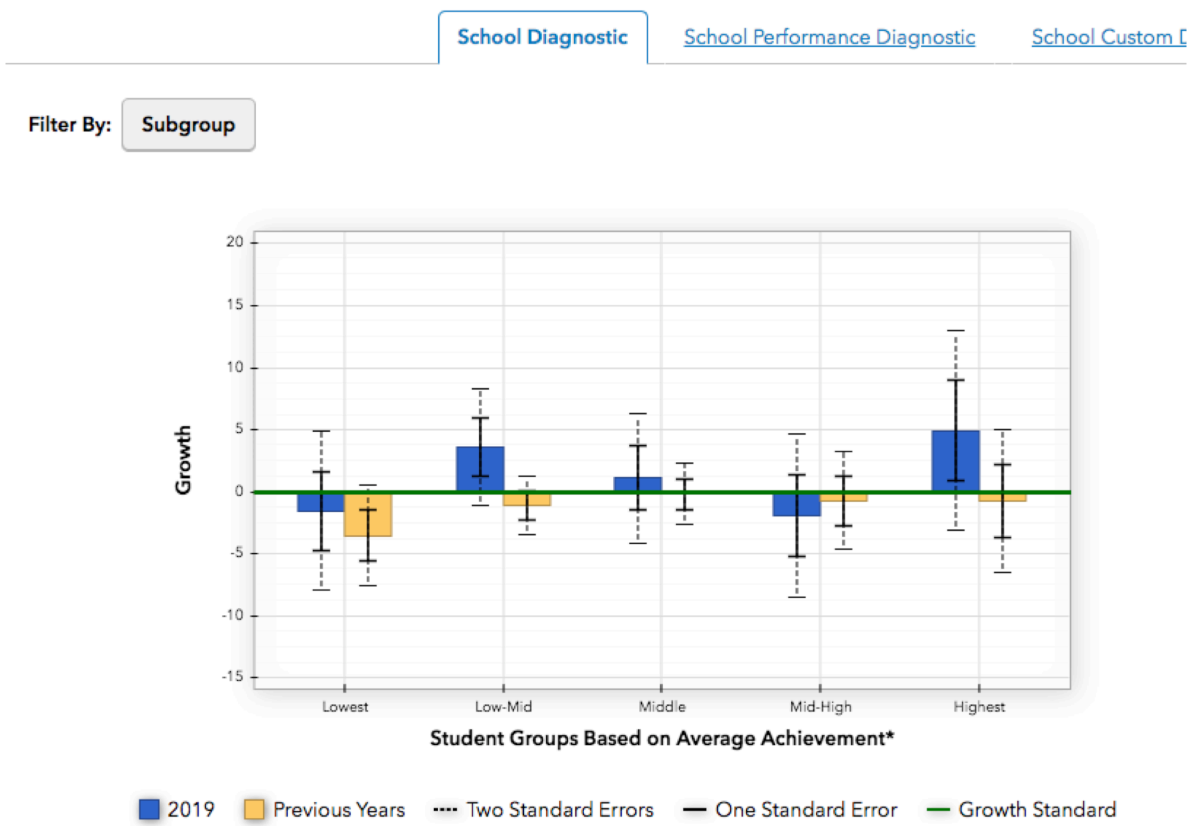
If special education students are disadvantaging a teacher's ability to receive a bonus, will the district have to re-consider the clustering of SWD and ELL students that allows for inclusion services to be rendered? Implications for the LEA will involve having discussions among district stakeholders concerning the way students are assigned to classes. The superintendent of the LEA where this study was conducted often says that leaders in the education community should be making decisions based on student services, not staff convenience. In regard to this principle, the curriculum and student service professionals, along with principals, teachers, and parents, must have dialogue about what learning scenario is best for ***students*** when assigning students to teachers. However, this discussion will have to be conducted also with an eye to the merit-based pay initiatives the state has established for teacher compensation. If principal-agent theory is informing student-teacher assignment practices based on what is assumed to be best for students, would this practice disadvantage a teacher's opportunity to earn a merit-based bonus because she happens to be effective at teaching special education students (Holmstrom & Milgrom, 1991)? Although she may be the best teacher for this group of students, the odds of her receiving the bonus are compromised when she has this group of SWD students in her classroom.

Another local concern that may manifest when bonuses upward of $7000 are at stake is that teachers may move to grade levels and schools where they feel opportunities for pay bonuses are most likely. District leaders will have to be cautious about honoring the requests of effective teachers to move out of primary grade classrooms (Kindergarten – 2$^{nd}$) where state and local bonuses are not available, to upper elementary grades or middle school classrooms where bonuses are available. This top-heavy system may severely handicap a school system's responsibility to have students reading on grade level before they come to the "tested grades" (Grades 3-8). If teachers were to internalize this research, I do not predict a danger of their looking to move between schools since all schools in the system have students with disabilities, the group that I have chronicled is correlated to classrooms where teachers did not receive math and reading bonuses. In addition, all schools have gifted students that are negatively correlated with teacher math bonuses. In fact, reason might suppose that teachers may look to leave higher SES communities where AIG student classification is more prevalent. Again, since we know that teachers are not solely driven by monetary rewards, I would not expect great migrations of teachers between schools although there may be some in-school grade level change requests (Guskey, 2017; Player, 2010).

After learning of the results of this research, what type of trainings will the LEA offer to its teachers? Will there be further differentiation training focused on the learning needs of its SWD subgroups based on instructional strategies in math and reading? Will they offer training on classroom management practices that instructors must master to maximize instructional minutes for those teachers whose EVAAS growth index is not meeting at least the state average, 0.0? Will the LEA leadership choose to dive deeper into the EVAAS instrument, focusing on the reports offered there to reflect on teaching practices that must be modified based on the diagnostic reports available to the grade level and individual teacher levels. The diagnostic report in table 5.2 shows a grade level report that is illustrating quintile student groups rather than teacher-level reports that are reported in tertile groupings.

Table 5.2

*School Quintile Growth Table for the 2017-2018 and 2018-2019 School Years*




* *Excerpted from EVAAS ®SAS diagnostic school level reporting*

The blue bars (first of each pair) indicate data from the most recent year, 2018-2019 student data, and the gold (the second of each pair), student and grade level growth data from 2017-2018, the year the research was completed. Reading the 2018-2019 data from this particular school, the most substantial growth occurred with the students that were predicted to perform in the $80^{th}$ to $99^{th}$ percentile and the $20^{th}$ to $40^{th}$ percentiles. The least student growth occurred in the $0^{th}$ to $20^{th}$ percentile and the $60^{th}$ to $80^{th}$ percentile student groups. The 0.0 horizontal mid-line on the table represents the state value added growth measure. In other words, this line marks the level where 50% of the teachers finish with a growth measure above this line and 50% of teachers below. Although this table strictly reflects grade level quintiles and not student characteristics, the diagnostic tables in EVAAS can be constructed to reflect quintile growth by student characteristic groups as well. In creating Table 5.2, I chose all students in a particular school and grade level so that there would be data sufficient to reflect growth indices in each quintile (a minimum of five student scores is needed to populate each quintile).

Sanders created EVAAS with the idea that it would provide for teacher reports enabling them to reflect on their teaching practices. These diagnostic tables do just that. At the grade level, subject area, and teacher levels, these reports allow teachers and principals to see which groups of students they were most successful in serving, and then to evaluate how to lesson plan in the current year to appropriately serve the learning needs of each student group. Although I did not select student groups down to the student characteristic level in Table 5.2, this option is available to teachers in the EVAAS software. Teachers also have the capability to access reports in EVAAS that illustrate which student groups grew the most and which grew the least as reflected by the value-added measure. I imagine that Sanders would have relished the opportunity to be involved in the professional learning community data meetings where these reports were created and discussed and student learning needs dissected in the search for improved instructional practices.

### Implications at the State Level

At the state level, if this study is indeed generalizable and repeatable, the implications should involve an intentional review of the merit-based pay system adopted by the North Carolina General Assembly. Although the goal of rewarding the most effective teachers is laudable, is the method the state is using to identify those teachers fair and focused on what is best for students (Amrein-Beardsley, 2014)? The bulk of current VAM research is critical of VAM models like EVAAS. However, despite these research findings, do the ends (teacher effectiveness) justify the means (using the EVAAS tool to evaluate teacher impact), if it can be found that student learning has been positively impacted by the incentives the bonuses are providing to North Carolina's teachers? Another goal of the bonus pay legislation that lawmakers must review is how effective the merit pay plan has been at attracting and keeping teacher talent. What positive impacts are these bonuses having on teacher retention and on attracting better professionals to the business of public education? A cost-benefit analysis where student results are measured against the cost of the merit-based program may be warranted at the state level to see if the state is truly getting appropriate results for the monetary investment of taxpayer dollars.

### Implications for Higher Learning

What implications does this study present to higher learning? What role do institutions of higher learning have in understanding how students grow, knowing what student groups are more likely to grow, and then using this knowledge to work with teacher trainees to leverage their instructional talents to maximize all student-learning outcomes? What responsibility do colleges and university researcher/instructors have in shaping public school curriculum and in instructing our future teachers to be the most impactful toward their students' learning?  The implications to the state of North Carolina concerning the findings in this research are wholly dependent on how it may be repeated on a macro-level.  If the findings of a state-level study were to yield similar results, especially with the *SWD* subgroup, the implications may be great on how teacher merit-based pay is structured in North Carolina.

**Implications for the Classroom**

The implication of this research on the classroom may be the most important revelation that comes from this research.  Most all research on the topic of student achievement cites the classroom teacher as the most important factor, that the school can control for, in determining student growth and achievement  (Sanders, 2003).  The teacher effect is more important than climate, principal effectiveness, class size, or curriculum (Sanders, Wright, & Horn, 1997; Teachers Matter, 2012).  EVAAS creates for teachers individual diagnostic reports that illustrate how effective teachers are at serving the tertiles of learners in a teacher's classroom, and the different subgroups of students in those classrooms (Table 5.2).  How will the classroom teacher use this personalized EVAAS reporting data to inform their instructional strategies going forward based on the findings in this research?  An understanding of instructional strategies, student learning-styles, and differentiated learning must be part of the professional development offerings for teachers that serve students with learning disabilities.  I would hope that good teachers with classroom experience do not request that students with disabilities be excluded from their classrooms.  However, with the incentives of the merit-based pay legislation, these students seem to disadvantage teachers' likelihood of receiving the bonus.  Principals, district administrators, and teachers must work together to do what is best for students while keeping in mind the implications of merit-based pay for effective teachers in the affected grades. The problem of student to teacher classroom assignment is a complex one.

**Recommendations for Future Research**

Logistical regression models tend to ignore collinearity issues amongst variables. Although outside the scope of this exploratory research, the relationship between *ELL* and *SES* variables are great, as is the relationship between the *SWD* and *White* predictor variables. Both these sets of variables' relationships would prove to have high levels of collinearity that might be meted out if these variables were regressed independently against the dependent variables. An understanding of how these predictor variables are correlated to the dependents without the collinearity issues associated with some logistic regression models would be beneficial in a future study.

In light of North Carolina's merit-based teacher compensation legislation, a study like this one that encompasses data from all across the state of North Carolina is important to conduct. Just as school systems and private industries undergo exhaustive program evaluations to discover if they are getting the results they desired after implementing a particular program, the state of North Carolina, if it is truly interested to see if the merit-based pay program is fair and valid, should undergo a macro-level study with a similar methodology to the one conducted in the local LEA I researched.

In a parallel setting, I would be interested in seeing the results of a similar study from other regions of the state to see what predictor variables they find to be significant and impactful in predicting teacher merit pay opportunities. As mentioned earlier, repeating this study where geography, population, culture, and socio-economic factors are different from the rural western North Carolina county where this study was completed would be valuable to policy-makers.  If this same study were completed a year from now in the same LEA, would the results be the same?  If not, what implications would this have for the LEA, for the EVAAS model and merit-based pay systems in general?

Finally, I think there is value in doing a mixed methods study that combines interviews and surveys of principals to find out their student-to-teacher assignment practices married with the exploratory quantitative research methodology used in this study.  Do principals "reward" staff with students more likely to assist their teachers in earning the bonus?  Do they use other student-to-teacher assignment practices designed to match teacher strengths and student needs?  The qualitative component of that study may go far to inform the procedures and policies a school system should consider going forward when it comes to assigning students to classrooms.

**Concluding Thoughts**

I continue to have trepidation over how the EVAAS tool is being used by school and district administrators to make high stakes decisions in North Carolina and other states where the VAM model is used.  If teachers have questions about the validity of the model and the high stakes decisions it is being used for, their concerns are valid.  Since most research finds almost no link between a teacher's degrees earned and the performance of the students in her classroom, the state discontinued the practice of compensating teachers for advanced degrees for those earned after 2016.  This gave the state a degree of flexibility to put more money into merit-based pay.  Much of the research referenced in this study though casts a great cloud of doubt over the inherent fairness of VAM models to make high stakes decisions.  Is there a way to triangulate the findings of exploratory quantitative studies like this one and qualitative study findings about principal student assignment practices to develop a pay plan that encourages and rewards effective teaching while allowing for the disadvantages that certain student groups may cause for teacher EVAAS scores?

Can more complex models that combine principal evaluations, student growth (EVAAS), and student characteristics be developed so that no teacher is disadvantaged by the communities where they serve, the schools where they work, and the children that they teach. I hope that the state of North Carolina can address the issues raised by my research. The EVAAS tool offers some valuable information to principals and teachers about how to improve their teaching. The idea of merit-based pay appeals to me as I am sure it does most tax payers who would agree that it is a worthwhile endeavor to create policy that rewards those teachers that most positively effect student learning. However, EVAAS was not designed for these high stakes decisions like teacher compensation and the problems this study has revealed are a testament to the concerns of its creator.

**Legal Issues**

From a legal standpoint, if this research is indeed generalizable, the state of North Carolina may have to face challenges in court based on property rights and due process charges. In Trout v. Knox County Board of Education, the plaintiffs argued that TVAAS was arbitrary and capricious and too imprecise to be used to assess teacher effectiveness and deny them a bonus (Paige, Amrein-Beardsley, & Close, 2019). The federal district court was sympathetic to the plaintiffs' charge insofar as the VAM they claimed was marked with statistical inadequacies. However, the court ultimately ruled on the side of the government over the claims of the plaintiffs (Paige, Amrein-Beardsley, & Close, 2019). On closer inspection, this case did not have the statistical specificity that my research provides for North Carolina teachers. Because the bonuses are applied to so few of North Carolina's public school teachers, and because many North Carolina public school teachers teach outside the merit pay subject areas, and because of the high stakes nature of the merit-based system, I foresee a challenge by the North Carolina Association of Educators or even by individual teachers in a state court over what they perceive as an unfair allocation of pay based on a flawed VAM model, EVAAS.

**Ethical Issues**

Measuring teacher efficacy is a complex issue, and an ethical one, that does not present an easy solution. How should the state honor the hard work of its teachers without a system that objectively measures the impact they have on their students' learning? Notwithstanding the merit-based pay problems this study has revealed, school systems will need to investigate inclusion and "pull out" models of special education instruction. Ethically, it is paramount that schools find which instructional models offer students the best chances to be successful, and that offer their teachers the best opportunity to show the growth necessary to receive merit-based pay.

**Questions Unanswered**

I have several unanswered questions about what the teacher accountability growth estimate or index is actually measuring. Since it is being used to evaluate effective teachers worthy of bonuses, answers to these questions are paramount to understanding the validity of the merit pay initiatives in the state. Although many of these questions fall outside the scope of my research, these queries may inform future mixed method research concerning the validity of the EVAAS measurement tool in North Carolina and other places the SAS tool is utilized. This series of questions raises concerns and identifies weaknesses inherent in the EVAAS model.

- Is the EVAAS teacher index measuring a teacher's impact on her student's learning or the impact of classroom peer effects on that child's learning?
- Is the index measuring the teacher's effectiveness or the principal's non-random student assignment practices?
- Is the growth estimate a legitimate reflection of teacher impact or a manifestation of the ceiling effect?
- Is the EVAAS index a representation of a teacher's effectiveness or the result of her serving the special education cluster in her grade level?
- Is EVAAS measuring teacher impact or the impact of the tutor that works three nights a week with the student?
- Is the teacher accountability growth measure evaluating a teacher's effect on student learning or the recent life change the child has suffered during her parents' divorce?

In addition to the measurement issues raised by these questions, the reality of a teacher's EVAAS growth index calculation is that it is not capable of separating out what part of the child's learning a teacher is responsible for when an effective and ineffective teacher both provide reading or math instruction to a child (spillage). However, when I consider my reservations toward placing worth in the VAM instrument, I remember the words of Winston Churchill and what he had to say about democracy. The English prime minister offered for consideration the idea that democracy was the worst form of government except for all the others. As an evaluation tool of teacher effectiveness, EVAAS may be the most objective tool we currently have in education to evaluate teacher impact. In other words, it may be the worst tool we have to measure teacher effectiveness… except for all the others.

Despite my cognitive wavering about the use of EVAAS for high stakes decisions like merit pay, and although it may be the best tool we currently have to measure teacher effectiveness, I cannot say that it is the fairest method by which to assign teacher bonuses based on the findings in this research.

## References


*The ABCs of Public Education* [Brochure]. (2006). Retrieved, from http://www.ncpublicschools.org/docs/accountability/reporting/abc/2005-06/abcsbrochure.pdf

AERA [American Educational Research Association]. (2015). AERA Statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, *44*(8), 448–452, DOI: 10.3102/0013189X15618385

Aldrich, M. W. (2017). William Sanders, pioneer of controversial value-added model for judging teachers, dies. *Chalkbeat*. Retrieved from https://www.chalkbeat.org/posts/tn/2017/03/20/william-sanders-pioneer-of-controversial-value-added-model-for-judging-teachers-dies/

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS EVAAS) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives,20*(12), 1-28. doi:10.14507/epaa.v20n12.2012

Amrein-Beardsley, A. (2014). Rethinking Value-Added Models in Education: Critical Perspectives on Tests and Assessment-Based Accountability. United Kingdom: Taylor & Francis.

Amrein-Bearsley, A., Pivovarova, M., & Geiger, T. J. (2016). Value-added models: What the experts say. *Phi Delta Kappan (98),* 35-40. Retrieved from https://www.kappanonline.org/value-added-models-what-the-experts-say/

ASA [American Statistical Association]. (2014). ASA statement on using value-added models for educational assessment. Retrieved from https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Babbie, E. R. *The practice of social research, 12th ed.* (2010). Belmont, CA: Wadsworth Cengage.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., . . . Shepard, L. A. (2010, August 29). *EPI Briefing Paper* (Rep. No. 278). Retrieved from Economic Policy Institute website: https://files.eric.ed.gov/fulltext/ED516803.pdf

Ballou, D., & Podgursky, M. (1998). Teacher recruitment and retention in public and private schools. *Journal of Policy Analysis and Management, 17*(3), 393–417.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-65. Retrieved from http://www.jstor.org.proxy006.nclive.org/stable/3701306

Ballou, D., & Springer, M. (2015). Using Student Test Scores to Measure Teacher Performance: Some Problems in the Design and Implementation of Evaluation Systems. *Educational Researcher*, *44*(2), 77. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=edsjsr&AN=edsjsr.24571534&site=eds-live&scope=site

Bandeira de Mello, V., Blankenship, C., & McLaughlin, D.H. (2009). Mapping state proficiency standards onto NAEP scales: 2005-2007 (NCES 2010-456). National Center for Education Statistics,Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/studies/2010456.pdf

Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review, 95*(3), 616-636. Retrieved from http://econ.lse. ac.uk/staff/mghatak/motivated.pdf

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Education Measurement: Issues and Practice, 28*(4), 42-51. Retrieved from https://www.nciea.org/publications/normative_criterion_growth_DB08.pdf

Blackstone, A. (2012). Sampling in quantitative research. In *Principles of sociological inquiry: Qualitative and quantitative methods* (1st ed.). Retrieved from https://saylordotorg.github.io/text_principles-of-sociological-inquiry-qualitative-and-quantitative-methods/s10-03-sampling-in-quantitative-resea.html

Braun, H. (2005). Using student progress to evaluate teachers: A primer on value-added models. *Policy Information PERSPECTIVE: Educational Testing Service,* 1-16. Retrieved from http://www.ets.org/Media/Research/pdf/PICVAM.pdf

Brophy, J. E. (1973). Stability of teacher effectiveness. *American Educational Research Journal, 10*(3), 245–252. doi:10.3102/00028312010003245

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, *31*(1), 51–82. Retrieved from https://doi-org.proxy006.nclive.org/http://www.journals.uchicago.edu/loi/jole

Burns, R. B., & Mason, D. A. (1995). Organizational constraints on the formation of elementary school classes. *American Journal of Education*, *103*(2), 185–212. doi:10.1086/444096

Carey, K. (2004). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *The Education Trust 8*(1), 1-44. Retrieved from https://edtrust.org/resource/the-real-value-of-teachers-using-new-information-about-teacher-effectiveness-to-close-the-achievement-gap/

Carnoy, M., & Garcia, E. (2017). Five key trends in U.S. student performance: Progress by blacks and Hispanics, the takeoff of Asians, the stall of non-English speakers, the persistence of socioeconomic gaps, and the damaging effect of highly segregated schools. *Economic Policy Institute*, 1–61. Retrieved from https://www.epi.org/files/pdf/113217.pdf

Cody, C. A., McFarland, J., Moore, J. E., & Preston, J. (2010). The evolution & use of growth models. *North Carolina Department of Public Instruction, Financial and Business Services Internship Program 4*(2), 1-12. Retrieved from http://www.ncpublicschools.org/docs/intern-research/reports/growth.pdf

Cohen, D. K., & Murnane, R. J. (1985). The merits of merit pay. *U.S. Department of Education: Office of Educational Research and Improvement*. Retrieved from https://files.eric.ed.gov/fulltext/ED270843.pdf

Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers college record, 116*(1) 1-34. Retrieved from http://www.tcrecord.org/Content.asp?ContentId=17291

Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice.* Providence, R.I.: Annenberg Institute for School Reform.

Creswell, J. W. (2005). *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Pearson.

Darling-Hammond, L., Amrein-Beardsley, A., & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan, 93*(6), 8-15. Retrieved from http://www.jstor.org.proxy006.nclive.org/stable/41497541

Dee, T., & Jacob, B. (2010). Evaluating NCLB. *Education Next, 10*(3). Retrieved from https://www.educationnext.org/evaluating-nclb/

Dills, A. K., & Mulholland, S. E. (2010). A comparative look at private and public schools' class size determinants. *Education Economics,18*(4), 435-454. doi:10.1080/09645290903546397

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources, 37*(4), 696–727.

Eckert, J. M., & Dabrowski, J. (2010). Should value-added measures be used for performance pay? *Phi Delta Kappan,91*(8), 88-92. Retrieved from http://eds.a.ebscohost.com.proxy006.nclive.org/eds/pdfviewer/pdfviewer?vid=4&sid=5a37a25e-8569-45af-8cf8-b66745787e83@sessionmgr4008

Field, A. P. (n.d.). Chapter 19: Logistic regression. In *Discovering statistics using SPSS*. (pp. 1-27). Retrieved from https://edge.sagepub.com/system/files/chapter19_3.pdf

Fiscal Year 2018-2019: North Carolina public school salary schedules. (n.d.). *Financial & Business Services: NC Department of Public Instruction*. Retrieved from http://www.ncpublicschools.org/docs/fbs/finance/salary/schedules/2018-19schedules.pdf

Foltz, B. (Producer). (2015, March 8). Statistics 101: Logistic regression, an introduction. Retrieved from www.youtube.com/watch?v=zAULhNrnuL4&list=PLIeGtxpvyG-JmBQ9XoFD4rs-b3hkcX7Uu

General Assembly of North Carolina. N.C. Gen. Stat. § 115C 335. 1998 5, s. 4. Raleigh, NC. 1998. Retrieved from http://www.ncga.state.nc.us/enactedlegislation/statutes/html/bysection/chapter_115c/gs_115c-335.html

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher,44*(2), 96-104.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher,44*(2), 96-104.

Gray, K. 1993. Why we will lose: Taylorism in America’s high schools. *The Phi Delta Kappan, 74*(5). http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=edsjsr&AN=edsjsr.20404887&site=eds-live&scope=site

Guba, E. G., & Lincoln, Y. S. (2008). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research, 3rd ed.* (pp. 255–286). Thousand Oaks, CA: Sage Publications, Inc. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=psyh&AN=2008-06349-008&site=eds-live&scope=site

Guskey, T. R. (2017). Where do you want to get to? *Learning Professional*, *38*(2), 32–37. Retrieved from https://login.proxy006.nclive.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=122515440&site=eds-live&scope=site

Guthrie, J. W., & Ray, J. R. (2008). *Modern school business administration : a planning approach*. Boston : Pearson Allyn and Bacon, 2008. Retrieved from https://login.proxy006.nclive.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat04042a&AN=app.b3003312&site=eds-live&scope=site

Guthrie, J. W., Springer, M. G., Rolle, R. A., & Houck, E. A. (2007). *Modern education finance and policy*. Mahwah, NJ: Allyn & Bacon.

Haertel, E. (2013). Reliability and validity of inferences about teachers based on student test scores. *ETS Research & Development Center for Research on Human Capital and Education.* Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/PICANG14.pdf

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal, 113*(485), F64–F98. doi:10.1111/1468-0297.00099

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2001). *Does Peer Ability Affect Student Achievement?* Retrieved from https://login.proxy006.nclive.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED476941&site=eds-live&scope=site

Harris, D. N. (2010). Clear away the smoke and mirrors of value-added. *Phi Delta Kappan,91*(8), 66-69. Retrieved from http://www.wiroundtable.org/Web_Site_PDFs/2010_pk12_committee/Harris PDK VA Final.pdf.

Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management, 29*(1), 183–208.

Helms, A. D. (2017, January 12). Berger to NC teacher group: Merit bonuses are the opposite of assembly line treatment. *Charlotte Observer*. Retrieved from https://www.charlotteobserver.com/news/local/education/article126074714.html

Hermann, M., Walsh, E., & Isenberg, E. (2016). Shrinkage of value-added estimates and characteristics of students with hard-to-predict achievement levels. *Statistics and Public Policy*, *3*(1), 1–10. htttps://doi.org/10.1080/2330443X.2016.1182878

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics & Organization*, *Special Issue* (24). Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=edshol&AN=edshol.hein.journals.jleo7.31&site=eds-live&scope=site

Horn, J., & Wilburn, D. (2013). *The mismeasure of education*. Charlotte, NC: Information Age Publishing.

Horne, J., Foley, V. P., & Flora, B. H. (2014). Race to the paycheck: Merit pay and theories of teacher motivation. *Journal of Academic Administration in Higher Education, 10*(1), 35-39. Retrieved from https://files.eric.ed.gov/fulltext/EJ1140899.pdf

Ireh, M. (2016, June). *Scientific management still endures in education* [Scholarly project]. Retrieved from https://files.eric.ed.gov/fulltext/ED566616.pdf

Jensen, E. (2009). *Teaching with poverty in mind. [electronic resource]: what being poor does to kids' brains and what schools can do about it*. ASCD. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=cat04042a&AN=app.b5570141&site=eds-live&scope=site

Jensen, M. C. (1983). Organization theory and methodology. *Accounting Review, 58*(2), 1-36. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=94036##

Johnson, S. M., & Papay, J. P. (2010). Expecting too much of performance pay? A study of four school districts uncovers the messy reality of adding merit bonuses to the salary scale. *School Administrators, 67*(3), 22–27. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=edsgao&AN=edsgcl.222229053&site=eds-live&scope=site

Kellough, J. E., & Lu, H. (1993). The paradox of merit pay in the public sector. *Review of Public Personnel Administration, 13(2),* 45-64. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=edsbig&AN=edsbig.A14087090&site=eds-live&scope=site

Klein, A. (2015, April 10). No child left behind: An overview. *Education Week.* Retrieved from https://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html

Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review, 28*(6), 682-692. doi: 10.1016/j.econedurev.2009.02.003

Kupermintz, H., Shepard, L., & Linn, R. (2001). *Teacher Effects as a Measure of Teacher Effectiveness: Construct Validity Considerations in TVAAS (Tennessee Value Added Assessment System)*. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=eric&AN=ED458295&site=eds-live&scope=site

Lachlan-Hache, L., & Castro, M. (2015). Proficiency or growth? An exploration of two approaches for writing student learning targets. *Performance Management Advantage: Evaluation & Professional Growth,* 1-12. Retrieved from https://www.air.org/sites/default/files/Exploration-of-Two-Approaches-Student-Learning-Targets-April-2015.pdf

Lockwood, J. R., and McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics*, (1) 223-252. doi:10.1214/07-EJS057

Logistic regression with Stata Chapter 1: Introduction to logistic regression with Stata. (2019). UCLA Institute for Digital Research & Education: University of California Regents. Retrieved from https://stats.idre.ucla.edu/stata/webbooks/logistic/chapter1/logistic-regression-with-statachapter-1-introduction-to-logistic-regression-with-stata/

Manna, P. (2006). *School's in : Federalism and the national education agenda*. Washington, D.C.: Georgetown University Press. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=e000xna&AN=217189&site=eds-live&scope=site

Marsden, D., & Richardson, R. (1994). Performing for pay? The effects of 'merit pay' on motivation in public service. *British Journal of Industrial Relations, 32*(2), 243-261. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=edsbl&AN=EN016522259&site=eds-live&scope=site

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mayer, D. P., Mullens, J. E., & Moore, M. T. (2000). Monitoring school quality: An indicators report, NCES 2001– 030. Washington, DC: U.S. Department of Education, National Center for Education Statistics. John Ralph, Project Officer. Retrieved from nces.ed.gov/pubs2001/2001030.pdf

McCaffrey, D. F., Han, B. & Lockwood, J. R. (2008, February). From data to bonuses: A case study of the issues related to awarding teachers pay on the basis of the students' progress. Paper presented at the conference on performance incentives: Their growing impact on American K-12 education. National Center on Performance Incentives at Vanderbilt University's Peabody College.

McColl, L. (n.d.). Analysing categorical survey data [Web log post]. Retrieved from https://select-statistics.co.uk/blog/analysing-categorical-survey-data/

Mertler, C. A., & Charles, C. M. (2005). *Introduction to educational research*. (7th ed.) Pearson. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=edshtl&AN=mdp.39015059577380&site=eds-live&scope=site

Meyer, R. & Dokumaci E. (2010). Value-added models and the next generation of assessments [Exploratory Seminar]. *Center for K-12 Assessment & Performance Management*. Retrieved from http://www.k12center.org/rsc/pdf/MeyerDokumaci PresenterSession4.pdf

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

Monk, D. (1987). Assigning elementary pupils to their teachers. *The Elementary School Journal,88*(2), 167-187. Retrieved from http://www.jstor.org.proxy006.nclive.org/stable/1002040

Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives,18*(23). Retrieved from http://epaa.asu.edu/ojs/article/view/810

North Carolina Department of Public Instruction. (2001). ABC program information: 2000-2001 – History of the ABCs program. Retrieved from http://www.dpi.state.nc.us/accountability/reporting/abc/2000-01/history

North Carolina Department of Public Instruction (2015). North Carolina Race to the Top closeout report 2015: Executive summary. 1-24. Retrieved July from https://www2.ed.gov/programs/racetothetop/state-reported-sharing/ncexsumm.pdf

North Carolina General Assembly, NC Senate. (n.d.). Teacher bonuses 2017 4th and 5th reading, 4th-8th math. Retrieved from http://www.ncleg.net/gascripts/BillLookUp/BillLookUp.pl?Session=2017&BillID=s257

North Carolina No Child Left Behind: Elementary and Secondary /Education Act (ESEA of 1965). (2010). Retrieved from http://www.ncpublicschools.org/docs/program-monitoring/titleIA/distinguished/2010.pdf

Oakes, M. M., & Robertson, J. S. (2014). North Carolina public school teachers: Reactions to teacher evaluations and merit pay. Retrieved from http://people.uncw.edu/robertsonj/documents/NC Teacher Evaluation and Merit Pay survey results Report.pdf

Paige, M. A., Amrein-Beardsley, A., & Close, K. (2019). Tennessee's National Impact on Teacher Evaluation Law & Policy: An Assessment of Value-Added Model Litigation. *Tennessee Journal of Law & Policy*, *13*(2), 523–574. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=a9h&AN=134919799&site=eds-live&scope=site

Paufler, N. A., & Amrein-Beardsley, A. (2014). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal (AERJ),51*(2), 328-362. http://dx.doi.org/10.3102/0002831213508299

Peng, C., Lee, K., & Ingersoll, G. (2002). An Introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, *96*(1), 3-14. doi:/10.1080/00220670209598786

Pham, L. D., Nguyen, T. D., & Springer, M. (2017, April). Teacher merit pay and student test scores: A meta-analysis[Scholarly project]. In *ResearchGate*. doi:10.13140/RG.2.2.34225.68961

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, Maryland: Rowman & Littlefield Publishers. Retrieved from http://search.ebscohost.com.proxy006.nclive.org/login.aspx?direct=true&db=cat04042a&AN=app.b2138524&site=eds-live&scope=site

Pitre-Martin, M. (2018). *North Carolina test coordinators' policies and procedures handbook.* Public Schools of North Carolina, Division of Accountability Services: North Carolina Testing Program. Raleigh, NC: State Board of Education: Department of Public Instruction.

Player, D. (2010). Non-monetary compensation in the public teacher labor market. *Education Finance and Policy, 5*(1), 82-103. doi:10.1162/edfp.2009.5.1.5105

Plymouth State University. (2019). Why is the IRB important? Retrieved from https://campus.plymouth.edu/institutional-review-board/why-irb/

Podgursky, M. J., & Springer, M. G. (2010). Market-and performance-based reforms of teacher compensation: A review of recent practices, policies, and research. Program on Education Policy and Governance at Harvard Kennedy School.

Reardon, S. F. (2013). The widening income academic gap. *Education Leadership, 70*(8), 10-16. Retrieved from http://www.ascd.org/publications/educational-leadership/may13/vol70/num08/The-Widening-Income-Achievement-Gap.aspx

Sacerdote, B. (2011). Peer effects in education: how might they work, how big are they and how much do we know thus far? In Hanushek, E., Machin, S., & Woessmann, L. (Eds.). *Handbook of the economics of education,* (1st ed.) (pp. 249-277). North Holland: Elsevier. doi: 10.1016/S0169-7218(11)03004-8

Sanders, W. L. (2003, April). Beyond "No Child Left Behind." Paper presented at the annual meeting of the American Educational Research Association, Chicago. Accessed at https://files.eric.ed.gov/fulltext/ED476149.pdf

Sanders, W., & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 12*(3), 247-256. Retrieved from https://www.sas.com/govedu/edu/ed_eval.pdf

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added system: A quantitative outcomes-based approach to educational assessment. In J. Millman (Ed.). *Grading teachers, grading schools: Is student achievement a valid measure?* (pp. 137-162). Thousand Oaks, CA: Corwin Press, Inc.

Sanders, W. L., Leandro, J. G., Rivers, J. C., & Wright, S. P. (2009). [White paper] A response to criticisms of SAS EVAAS. Retrieved from https://pdfs.semanticscholar.org/9eac/fc1e23f576d51ea71813a6c8d673caa65c5e.pdf?_ga=2.242429185.208933041.1498842929-987755573.1498319466

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57. doi-org.proxy006.nclive.org/10.1023/a:1007999204543

Social and behavioral responsible conduct of research course 1. (2019). [CITI Program]. Retrieved from https://www.citiprogram.org/members/index.cfm?pageID=50

Social/behavioral research course. (2019). [CITI Program]. Retrieved from https://www.citiprogram.org/members/index.cfm?pageID=50

Sorensen, T. B. (2016). Value-added measurement or modeling (VAM). *Education International,2.0*, 1-28. Retrieved from https://www.ei-ie.org/media_gallery/2016_EI_VAM_EN_final_Web.pdf.

Springer, M. G. (2009). *Performance Incentives : Their Growing Impact on American K-12 Education*. Washington, D.C.: Brookings Institution Press. Retrieved from https://login.proxy006.nclive.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=301813&site=eds-live&scope=site

Springer, M.G. & Taylor, L.L. (2016). Designing incentives for public school teachers: Evidence from a Texas incentive pay program. *Journal of Education Finance, 41*(3), 344-381.

Stoops, T. (2018, July). Testing and accountability. Raleigh, North Carolina: John Locke Foundation. Retrieved from https://www.johnlocke.org/policy-position/testing-and-accountability/

Stepwise regression. (n.d.). In *NCSS Statistical Software* (Chapter 311). Retrieved from https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Stepwise_Regression.pdf

Teacher bonuses –House Bill –HS IB, AP, CTE HB – 840. Retrieved from http://www.ncleg.net/Sessions/2017/Bills/House/PDF/H840v1.pdf

Teachers matter: Understanding teachers' impact on student achievement. (2012). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/corporate_pubs/CP693z1-2012-09.html

The facts on merit pay for teachers from Public Schools First NC. (2016, January 19). Retrieved from http://www.publicschoolsfirstnc.org/wp-content/uploads/2016/03/01-19-2016-Merit-Pay-for-Teachers.pdf

Thomas, P. L. (2013). Testing capitalism: Perpetuating privilege behind the masks of merit and objectivity. *International Education Journal: Comparative Perspectives*, *12*(2), 85–103. Retrieved from https://login.proxy006.nclive.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1017716&site=eds-live&scope=site

Wagner, L. (2014, March). Students, teachers grapple with Read to Achieve law. *NC Policy Watch*. Retrieved from http://www.ncpolicywatch.com/2014/03/20/students-teachers-grapple-with-read-to-achieve-law/

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

WestEd. (2012). [Working Paper]. Validating North Carolina's educator evaluation system: Evidence in chain of claims to support system use. Retrieved from http://www.ncpublicschools.org/docs/effectiveness-model/evaas/selection/validating-evaluation.pdf

Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2015). Getting classroom observations right: Lesson on how from four pioneering districts. *Education Next,15*(1). Retrieved from http://educationnext.org/getting-classroom-observations-right/

Wiley, E. W. (2006). *A practitioner's guide to value added assessment*. University of Colorado: Educational Policy Studies Laboratory Research Monograph. Retrieved from https://nepc.colorado.edu/sites/default/files/Wiley_APractitionersGuide.pdf

Wuensch, K. L. (2014). Binary logistic regression with SPSS. 1-29. Retrieved from http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.PDF

Yuan, K. (2015). A value-added study of teacher spillover effects across four core subjects in middle schools. *Education Policy Analysis Archives, 23*(38), 1–24.

## APPENDIX A
## THREE CATEGORIES OF GROWTH MODELS

**SIMPLE GROWTH (Similar to EVAAS multi-variate model for EOGs)**

**Function –** Measures difference in student's scale scores from year to year

**Usage –** Can be used to calculate school, district, and state growth

**GROWTH-TO-STANDARD**

**Function –** Shows if students are on track to meet proficiency standards

**Usage –** Can be used to calculate school, district, and state growth and proficiency

**PROJECTION (EVAAS univariate model for EOC and NCFE end of year exams)**

**Function –** Predicts students' academic levels based on previous test scores

**Usage –** Can be used to calculate growth and proficiency at the student, school, district, and state levels

* *Exerpted from The Evolution & Use of Growth Models*
Christopher A. Cody, Joel McFarland, J. Eric Morre, & Jennifer Preston, 2010

## APPENDIX B

**TEACHER EFFECTIVENESS**

**Per federal requirements, the State must adopt definitions of effective and highly effective teachers.**

A ***highly effective teacher*** is one who receives a rating of at least "accomplished" on each of the Teacher Evaluation Standards 1 - 5 and receives a rating of "exceeds expected growth" on Standard 6 of the Teacher Evaluation Instrument. The End-of-Course assessments, End-of-Grade assessments, Career and Technical Education Post-Assessments, and the Measures of Student Learning provide the student data used to calculate the growth value.

An ***effective teacher*** is one who receives a rating of at least "proficient" on each of the Teacher Evaluation Standards 1-5 and receives a rating of at least "meets expected growth" on Standard 6 of the Teacher Evaluation Instrument.

A teacher in ***need of improvement*** is one who fails to receive a rating of at least "proficient" on each of the Teacher Evaluation Standards 1-5 or receives a rating of "does not meet expected growth" on Standard 6 of the Teacher Evaluation Instrument.

A three-year rolling average of student growth values generates the sixth standard rating used to determine teacher effectiveness. Only student growth values based on the individual students taught by a teacher will be used to determine the three-year rolling average for that teacher.

** Excerpted from: http://www.ncpublicschools.org/docs/effectiveness-model/ncees/standards/prof-teach-standards.pdf*

**APPENDIX C**

**EVAAS Growth Indexes: Defining Teacher Effectiveness in terms of standard errors**

| **Value Added Color and Teacher Measure Designation** | **Growth Measure Compared to the Growth Standard** | **Index*** | **Interpretation** |
|---|---|---|---|
| **Exceeds Expected Growth** | **At least 2 standard errors above** | **2.00 or greater** | **Significant evidence that students made more progress than the Growth Standard** |
| **Meets Expected Growth** | **Between 2 standard errors above and 2 standard errors below** | **Between -2.00 and 2.00** | **Evidence that students made progress similar to the Growth Standard** |
| **Does Not Meet Expected Growth** | **More than 2 standard errors below** | **Less than -2.00** | **Significant evidence that students made less progress than the Growth Standard** |

Note: When an index falls exactly on the boundary between two colors, the higher growth color is assigned.
*These rules for effectiveness levels and growth colors apply to all index values in the district, school, and teacher reports

** *Excerpted from North Carolina Department of Public Instruction: Technical Documentation for 2016 EVAAS Analyses* https://ncdpi.sas.com/support/EVAAS-NC-TechnicalDocumentation-2016.pdf

## APPENDIX D

**2018-2019 North Carolina Bachelor's Degree Certified Teacher Salary Schedule (Effective July 1, 2018)**

| Years of Experience | Monthly Salary (10 months) | Annual Salary |
|---|---|---|
| 0 | $3500 | $35,000 |
| 1 | $3600 | $36,000 |
| 2 | $3700 | $37,000 |
| 3 | $3800 | $38,000 |
| 4 | $3900 | $39,000 |
| 5 | $4000 | $40,000 |
| 6 | $4100 | $41,000 |
| 7 | $4200 | $42,000 |
| 8 | $4300 | $43,000 |
| 9 | $4400 | $44,000 |
| 10 | $4500 | $45,000 |
| 11 | $4600 | $46,000 |
| 12 | $4700 | $47,000 |
| 13 | $4800 | $48,000 |
| 14 | $4900 | $49,000 |
| 15 - 24 | $5000 | $50,000 |
| 25+ | $5200 | $52,000 |

* Excerpted from *Fiscal Year 2018 - 2019 North Carolina Public School Salary Schedules* at www.ncpublicschools.org/fbs/finance/salary/ Financial & Business Services NC Department of Public Instruction

## APPENDIX E

**Binary Logistic Regression – Grades 3-5 Reading**
**Dependent Variable Bonus Coded as 1: (RDGB _1)**

### Case Processing Summary

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 2925 | 76.7 |
| | Missing Cases | 891 | 23.3 |
| | Total | 3816 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 3816 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

### Dependent Variable Encoding

| Original Value | Internal Value |
|---|---|
| .00 | 0 |
| 1.00 | 1 |

## Block 0: Beginning Block

### Classification Table[a,b]

| | Observed | | RDGB_1 Predicted .00 | 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 0 | RDGB_1 | .00 | 1578 | 0 | 100.0 |
| | | 1.00 | 1347 | 0 | .0 |
| | Overall Percentage | | | | 53.9 |

a. Constant is included in the model.
b. The cut value is .500

## Variables in the Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 0 | Constant | -.158 | .037 | 18.205 | 1 | .000 | .854 |

## Variables not in the Equation[a]

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | Asian | 1.963 | 1 | .161 |
| | | Black | 10.804 | 1 | .001 |
| | | Hispanic | 39.054 | 1 | .000 |
| | | Indian | .013 | 1 | .911 |
| | | Mixed | 1.859 | 1 | .173 |
| | | White | 41.138 | 1 | .000 |
| | | PacificIs | .013 | 1 | .911 |
| | | F | 1.286 | 1 | .257 |
| | | M | 1.286 | 1 | .257 |
| | | AIG | 4.096 | 1 | .043 |
| | | SWD | 8.526 | 1 | .004 |
| | | SES | 3.635 | 1 | .057 |
| | | ELL | 10.521 | 1 | .001 |

a. Residual Chi-Squares are not computed because of redundancies.

# Block 1: Method = Forward Stepwise (Conditional)

## Omnibus Tests of Model Coefficients

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 41.447 | 1 | .000 |
| | Block | 41.447 | 1 | .000 |
| | Model | 41.447 | 1 | .000 |
| Step 2 | Step | 15.507 | 1 | .000 |
| | Block | 56.954 | 2 | .000 |

| | | | | |
|---|---|---|---|---|
| | Model | 56.954 | 2 | .000 |
| Step 3 | Step | 9.059 | 1 | .003 |
| | Block | 66.013 | 3 | .000 |
| | Model | 66.013 | 3 | .000 |
| Step 4 | Step | 4.215 | 1 | .040 |
| | Block | 70.228 | 4 | .000 |
| | Model | 70.228 | 4 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 3995.202[a] | .014 | .019 |
| 2 | 3979.695[a] | .019 | .026 |
| 3 | 3970.636[a] | .022 | .030 |
| 4 | 3966.421[a] | .024 | .032 |

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | Observed | | Predicted RDGB_1 .00 | 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | RDGB_1 | .00 | 631 | 947 | 40.0 |
| | | 1.00 | 386 | 961 | 71.3 |
| | Overall Percentage | | | | 54.4 |
| Step 2 | RDGB_1 | .00 | 544 | 1034 | 34.5 |
| | | 1.00 | 295 | 1052 | 78.1 |
| | Overall Percentage | | | | 54.6 |

| | | | | | |
|---|---|---|---|---|---|
| Step 3 | RDGB_1 | .00 | 756 | 822 | 47.9 |
| | | 1.00 | 462 | 885 | 65.7 |
| | Overall Percentage | | | | 56.1 |
| Step 4 | RDGB_1 | .00 | 756 | 822 | 47.9 |
| | | 1.00 | 462 | 885 | 65.7 |
| | Overall Percentage | | | | 56.1 |

a. The cut value is .500

## Variables in the Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | White | .506 | .079 | 40.845 | 1 | .000 | 1.659 |
| | Constant | -.491 | .065 | 57.848 | 1 | .000 | .612 |
| Step 2[b] | Asian | .657 | .166 | 15.573 | 1 | .000 | 1.929 |
| | White | .627 | .086 | 53.612 | 1 | .000 | 1.871 |
| | Constant | -.612 | .072 | 71.635 | 1 | .000 | .542 |
| Step 3[c] | Asian | .642 | .167 | 14.825 | 1 | .000 | 1.900 |
| | White | .632 | .086 | 54.363 | 1 | .000 | 1.882 |
| | SWD | -.296 | .099 | 8.963 | 1 | .003 | .744 |
| | Constant | -.562 | .074 | 57.495 | 1 | .000 | .570 |
| Step 4[d] | Asian | .723 | .172 | 17.751 | 1 | .000 | 2.062 |
| | Mixed | .351 | .170 | 4.265 | 1 | .039 | 1.420 |
| | White | .714 | .095 | 56.327 | 1 | .000 | 2.042 |
| | SWD | -.301 | .099 | 9.245 | 1 | .002 | .740 |
| | Constant | -.643 | .085 | 57.765 | 1 | .000 | .526 |

a. Variable(s) entered on step 1: White.

b. Variable(s) entered on step 2: Asian.

c. Variable(s) entered on step 3: SWD.

d. Variable(s) entered on step 4: Mixed.

### Model if Term Removed[a]

| Variable | | Model Log Likelihood | Change in -2 Log Likelihood | df | Sig. of the Change |
|---|---|---|---|---|---|
| Step 1 | White | -2018.330 | 41.457 | 1 | .000 |
| Step 2 | Asian | -1997.602 | 15.510 | 1 | .000 |
| | White | -2017.362 | 55.030 | 1 | .000 |
| Step 3 | Asian | -1992.702 | 14.769 | 1 | .000 |
| | White | -2013.230 | 55.823 | 1 | .000 |
| | SWD | -1989.848 | 9.060 | 1 | .003 |
| Step 4 | Asian | -1992.066 | 17.711 | 1 | .000 |
| | Mixed | -1985.319 | 4.217 | 1 | .040 |
| | White | -2012.476 | 58.531 | 1 | .000 |
| | SWD | -1987.883 | 9.345 | 1 | .002 |

a. Based on conditional parameter estimates

### Variables not in the Equation[a]

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 1 | Variables | Asian | 15.889 | 1 | .000 |
| | | Black | 2.218 | 1 | .136 |
| | | Hispanic | 8.432 | 1 | .004 |
| | | Indian | .123 | 1 | .725 |
| | | Mixed | 1.083 | 1 | .298 |
| | | PacificIs | .123 | 1 | .725 |
| | | F | .816 | 1 | .366 |
| | | M | .816 | 1 | .366 |
| | | AIG | 1.287 | 1 | .257 |
| | | SWD | 9.732 | 1 | .002 |
| | | SES | .000 | 1 | .994 |
| | | ELL | .091 | 1 | .763 |

| | | | | | |
|---|---|---|---|---|---|
| Step 2 | Variables | Black | .654 | 1 | .419 |
| | | Hispanic | 1.439 | 1 | .230 |
| | | Indian | .194 | 1 | .660 |
| | | Mixed | 3.993 | 1 | .046 |
| | | PacificIs | .194 | 1 | .660 |
| | | F | 1.211 | 1 | .271 |
| | | M | 1.211 | 1 | .271 |
| | | AIG | 1.075 | 1 | .300 |
| | | SWD | 8.999 | 1 | .003 |
| | | SES | .082 | 1 | .774 |
| | | ELL | .316 | 1 | .574 |
| Step 3 | Variables | Black | .556 | 1 | .456 |
| | | Hispanic | 1.696 | 1 | .193 |
| | | Indian | .163 | 1 | .687 |
| | | Mixed | 4.287 | 1 | .038 |
| | | PacificIs | .163 | 1 | .687 |
| | | F | .549 | 1 | .459 |
| | | M | .549 | 1 | .459 |
| | | AIG | .252 | 1 | .615 |
| | | SES | .377 | 1 | .539 |
| | | ELL | .105 | 1 | .746 |
| Step 4 | Variables | Black | .077 | 1 | .782 |
| | | Hispanic | .021 | 1 | .884 |
| | | Indian | .215 | 1 | .643 |
| | | PacificIs | .215 | 1 | .643 |
| | | F | .642 | 1 | .423 |
| | | M | .642 | 1 | .423 |
| | | AIG | .199 | 1 | .656 |
| | | SES | .540 | 1 | .463 |
| | | ELL | .088 | 1 | .766 |

**APPENDIX F**
**Binary Logistic Regression – Grades 3-5 Reading**
**Dependent Variable Bonus Coded as 0: (RDGB _0)**

### Case Processing Summary

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 2925 | 76.7 |
| | Missing Cases | 891 | 23.3 |
| | Total | 3816 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 3816 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

### Dependent Variable Encoding

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

## Block 0: Beginning Block

### Classification Table[a,b]

| | Observed | | Predicted RDGB_0 0 | Predicted RDGB_0 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 0 | RDGB_0 | 0 | 0 | 1347 | .0 |
| | | 1 | 0 | 1578 | 100.0 |
| | Overall Percentage | | | | 53.9 |

a. Constant is included in the model.

b. The cut value is .500

### Variables in the Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 0 | Constant | .158 | .037 | 18.205 | 1 | .000 | 1.171 |

### Variables not in the Equation[a]

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | Asian | 1.963 | 1 | .161 |
| | | Black | 10.804 | 1 | .001 |
| | | Hispanic | 39.054 | 1 | .000 |
| | | Indian | .013 | 1 | .911 |
| | | Mixed | 1.859 | 1 | .173 |
| | | White | 41.138 | 1 | .000 |
| | | PacificIs | .013 | 1 | .911 |
| | | F | 1.286 | 1 | .257 |
| | | M | 1.286 | 1 | .257 |
| | | AIG | 4.096 | 1 | .043 |
| | | SWD | 8.526 | 1 | .004 |
| | | SES | 3.635 | 1 | .057 |
| | | ELL | 10.521 | 1 | .001 |

## Block 1: Method = Forward Stepwise (Conditional)

### Omnibus Tests of Model Coefficients

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 41.447 | 1 | .000 |
| | Block | 41.447 | 1 | .000 |
| | Model | 41.447 | 1 | .000 |
| Step 2 | Step | 15.507 | 1 | .000 |
| | Block | 56.954 | 2 | .000 |

| | | | | |
|---|---|---|---|---|
| | Model | 56.954 | 2 | .000 |
| Step 3 | Step | 9.059 | 1 | .003 |
| | Block | 66.013 | 3 | .000 |
| | Model | 66.013 | 3 | .000 |
| Step 4 | Step | 4.215 | 1 | .040 |
| | Block | 70.228 | 4 | .000 |
| | Model | 70.228 | 4 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 3995.202[a] | .014 | .019 |
| 2 | 3979.695[a] | .019 | .026 |
| 3 | 3970.636[a] | .022 | .030 |
| 4 | 3966.421[a] | .024 | .032 |

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | Observed | | Predicted RDGB_0 0 | 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | RDGB_0 | 0 | 961 | 386 | 71.3 |
| | | 1 | 947 | 631 | 40.0 |
| | Overall Percentage | | | | 54.4 |
| Step 2 | RDGB_0 | 0 | 1052 | 295 | 78.1 |
| | | 1 | 1034 | 544 | 34.5 |
| | Overall Percentage | | | | 54.6 |
| Step 3 | RDGB_0 | 0 | 885 | 462 | 65.7 |
| | | 1 | 822 | 756 | 47.9 |
| | Overall Percentage | | | | 56.1 |

| | | | | | |
|---|---|---|---|---|---|
| Step 4 | RDGB_0 | 0 | 885 | 462 | 65.7 |
| | | 1 | 822 | 756 | 47.9 |
| | Overall Percentage | | | | 56.1 |

a. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | White | -.506 | .079 | 40.845 | 1 | .000 | .603 |
| | Constant | .491 | .065 | 57.848 | 1 | .000 | 1.635 |
| Step 2[b] | Asian | -.657 | .166 | 15.573 | 1 | .000 | .518 |
| | White | -.627 | .086 | 53.612 | 1 | .000 | .534 |
| | Constant | .612 | .072 | 71.635 | 1 | .000 | 1.844 |
| Step 3[c] | Asian | -.642 | .167 | 14.825 | 1 | .000 | .526 |
| | White | -.632 | .086 | 54.363 | 1 | .000 | .531 |
| | SWD | .296 | .099 | 8.963 | 1 | .003 | 1.344 |
| | Constant | .562 | .074 | 57.495 | 1 | .000 | 1.754 |
| Step 4[d] | Asian | -.723 | .172 | 17.751 | 1 | .000 | .485 |
| | Mixed | -.351 | .170 | 4.265 | 1 | .039 | .704 |
| | White | -.714 | .095 | 56.327 | 1 | .000 | .490 |
| | SWD | .301 | .099 | 9.245 | 1 | .002 | 1.351 |
| | Constant | .643 | .085 | 57.765 | 1 | .000 | 1.902 |

a. Variable(s) entered on step 1: White.

b. Variable(s) entered on step 2: Asian.

c. Variable(s) entered on step 3: SWD.

d. Variable(s) entered on step 4: Mixed.

### Model if Term Removed[a]

| Variable | | Model Log Likelihood | Change in -2 Log Likelihood | df | Sig. of the Change |
|---|---|---|---|---|---|
| Step 1 | White | -2018.330 | 41.457 | 1 | .000 |
| Step 2 | Asian | -1997.602 | 15.510 | 1 | .000 |
| | White | -2017.362 | 55.030 | 1 | .000 |
| Step 3 | Asian | -1992.702 | 14.769 | 1 | .000 |
| | White | -2013.230 | 55.823 | 1 | .000 |
| | SWD | -1989.848 | 9.060 | 1 | .003 |
| Step 4 | Asian | -1992.066 | 17.711 | 1 | .000 |
| | Mixed | -1985.319 | 4.217 | 1 | .040 |
| | White | -2012.476 | 58.531 | 1 | .000 |
| | SWD | -1987.883 | 9.345 | 1 | .002 |

a. Based on conditional parameter estimates

### Variables not in the Equation[a]

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 1 | Variables | Asian | 15.889 | 1 | .000 |
| | | Black | 2.218 | 1 | .136 |
| | | Hispanic | 8.432 | 1 | .004 |
| | | Indian | .123 | 1 | .725 |
| | | Mixed | 1.083 | 1 | .298 |
| | | PacificIs | .123 | 1 | .725 |
| | | F | .816 | 1 | .366 |
| | | M | .816 | 1 | .366 |
| | | AIG | 1.287 | 1 | .257 |
| | | SWD | 9.732 | 1 | .002 |
| | | SES | .000 | 1 | .994 |
| | | ELL | .091 | 1 | .763 |

| | | | | | |
|---|---|---|---|---|---|
| Step 2 | Variables | Black | .654 | 1 | .419 |
| | | Hispanic | 1.439 | 1 | .230 |
| | | Indian | .194 | 1 | .660 |
| | | Mixed | 3.993 | 1 | .046 |
| | | PacificIs | .194 | 1 | .660 |
| | | F | 1.211 | 1 | .271 |
| | | M | 1.211 | 1 | .271 |
| | | AIG | 1.075 | 1 | .300 |
| | | SWD | 8.999 | 1 | .003 |
| | | SES | .082 | 1 | .774 |
| | | ELL | .316 | 1 | .574 |
| Step 3 | Variables | Black | .556 | 1 | .456 |
| | | Hispanic | 1.696 | 1 | .193 |
| | | Indian | .163 | 1 | .687 |
| | | Mixed | 4.287 | 1 | .038 |
| | | PacificIs | .163 | 1 | .687 |
| | | F | .549 | 1 | .459 |
| | | M | .549 | 1 | .459 |
| | | AIG | .252 | 1 | .615 |
| | | SES | .377 | 1 | .539 |
| | | ELL | .105 | 1 | .746 |
| Step 4 | Variables | Black | .077 | 1 | .782 |
| | | Hispanic | .021 | 1 | .884 |
| | | Indian | .215 | 1 | .643 |
| | | PacificIs | .215 | 1 | .643 |
| | | F | .642 | 1 | .423 |
| | | M | .642 | 1 | .423 |
| | | AIG | .199 | 1 | .656 |
| | | SES | .540 | 1 | .463 |
| | | ELL | .088 | 1 | .766 |

## Appendix G – Binary Logistic Regression – Grades 4-8 Math Dependent Variable Bonus Coded as 1: (MTHB _1)

### Logistic Regression

**Case Processing Summary**

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 4490 | 100.0 |
| | Missing Cases | 1 | .0 |
| | Total | 4491 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 4491 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

### Dependent Variable Encoding

| Original Value | Internal Value |
|---|---|
| .00 | 0 |
| 1.00 | 1 |

### Block 0: Beginning Block

**Classification Table[a,b]**

| | Observed | | Predicted MTHB_1 .00 | Predicted MTHB_1 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 0 | MTHB_1 | .00 | 2738 | 0 | 100.0 |
| | | 1.00 | 1752 | 0 | .0 |
| | Overall Percentage | | | | 61.0 |

a. Constant is included in the model.
b. The cut value is .500

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 0 | Constant | -.446 | .031 | 212.964 | 1 | .000 | .640 |

### Variables not in the Equation[a]

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | Asian | 1.335 | 1 | .248 |
| | | Black | 7.762 | 1 | .005 |
| | | Hispanic | 5.130 | 1 | .024 |
| | | Indian | .964 | 1 | .326 |
| | | Mixed | 2.414 | 1 | .120 |
| | | White | 10.520 | 1 | .001 |
| | | PacificIs | .322 | 1 | .571 |
| | | F | .113 | 1 | .737 |
| | | M | .113 | 1 | .737 |
| | | AIG | 6.612 | 1 | .010 |
| | | SWD | 13.993 | 1 | .000 |
| | | SES | 4.754 | 1 | .029 |
| | | ELL | .642 | 1 | .423 |

a. Residual Chi-Squares are not computed because of redundancies.

## Block 1: Method = Forward Stepwise (Conditional)

### Omnibus Tests of Model Coefficients

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 14.249 | 1 | .000 |
| | Block | 14.249 | 1 | .000 |
| | Model | 14.249 | 1 | .000 |
| Step 2 | Step | 11.577 | 1 | .001 |
| | Block | 25.826 | 2 | .000 |
| | Model | 25.826 | 2 | .000 |
| Step 3 | Step | 7.919 | 1 | .005 |
| | Block | 33.745 | 3 | .000 |
| | Model | 33.745 | 3 | .000 |

| | | | | |
|---|---|---|---|---|
| Step 4 | Step | 4.171 | 1 | .041 |
| | Block | 37.916 | 4 | .000 |
| | Model | 37.916 | 4 | .000 |
| Step 5 | Step | 3.867 | 1 | .049 |
| | Block | 41.783 | 5 | .000 |
| | Model | 41.783 | 5 | .000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 5991.913[a] | .003 | .004 |
| 2 | 5980.337[a] | .006 | .008 |
| 3 | 5972.417[a] | .007 | .010 |
| 4 | 5968.246[a] | .008 | .011 |
| 5 | 5964.379[a] | .009 | .013 |

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

**Classification Table[a]**

| | Observed | | Predicted MTHB_1 .00 | Predicted MTHB_1 1.00 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | MTHB_1 | .00 | 2738 | 0 | 100.0 |
| | | 1.00 | 1752 | 0 | .0 |
| | Overall Percentage | | | | 61.0 |
| Step 2 | MTHB_1 | .00 | 2738 | 0 | 100.0 |
| | | 1.00 | 1752 | 0 | .0 |
| | Overall Percentage | | | | 61.0 |
| Step 3 | MTHB_1 | .00 | 2670 | 68 | 97.5 |
| | | 1.00 | 1684 | 68 | 3.9 |
| | Overall Percentage | | | | 61.0 |

| | | | | | |
|---|---|---|---|---|---|
| Step 4 | MTHB_1 | .00 | 2670 | 68 | 97.5 |
| | | 1.00 | 1684 | 68 | 3.9 |
| | Overall Percentage | | | | 61.0 |
| Step 5 | MTHB_1 | .00 | 2670 | 68 | 97.5 |
| | | 1.00 | 1684 | 68 | 3.9 |
| | Overall Percentage | | | | 61.0 |

a. The cut value is .500

## Variables in the Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | SWD | -.320 | .086 | 13.918 | 1 | .000 | .727 |
| | Constant | -.396 | .033 | 142.121 | 1 | .000 | .673 |
| Step 2[b] | AIG | -.272 | .080 | 11.411 | 1 | .001 | .762 |
| | SWD | -.377 | .087 | 18.670 | 1 | .000 | .686 |
| | Constant | -.336 | .038 | 79.917 | 1 | .000 | .715 |
| Step 3[c] | Black | .408 | .144 | 7.993 | 1 | .005 | 1.504 |
| | AIG | -.259 | .081 | 10.368 | 1 | .001 | .771 |
| | SWD | -.390 | .088 | 19.823 | 1 | .000 | .677 |
| | Constant | -.356 | .038 | 86.346 | 1 | .000 | .701 |
| Step 4[d] | Black | .440 | .145 | 9.205 | 1 | .002 | 1.553 |
| | Hispanic | .166 | .081 | 4.195 | 1 | .041 | 1.181 |
| | AIG | -.238 | .081 | 8.571 | 1 | .003 | .788 |
| | SWD | -.381 | .088 | 18.877 | 1 | .000 | .683 |
| | Constant | -.392 | .042 | 85.967 | 1 | .000 | .676 |
| Step 5[e] | Black | .462 | .146 | 10.079 | 1 | .001 | 1.588 |
| | Hispanic | .188 | .082 | 5.255 | 1 | .022 | 1.207 |
| | Mixed | .248 | .125 | 3.908 | 1 | .048 | 1.281 |
| | AIG | -.231 | .081 | 8.036 | 1 | .005 | .794 |
| | SWD | -.382 | .088 | 18.927 | 1 | .000 | .683 |
| | Constant | -.414 | .044 | 89.377 | 1 | .000 | .661 |

a. Variable(s) entered on step 1: SWD.

b. Variable(s) entered on step 2: AIG.

c. Variable(s) entered on step 3: Black.

d. Variable(s) entered on step 4: Hispanic.

e. Variable(s) entered on step 5: Mixed.

**Model if Term Removed[a]**

| Variable | | Model Log Likelihood | Change in -2 Log Likelihood | df | Sig. of the Change |
|---|---|---|---|---|---|
| Step 1 | SWD | -3003.083 | 14.253 | 1 | .000 |
| Step 2 | AIG | -2995.958 | 11.579 | 1 | .001 |
| | SWD | -2999.746 | 19.155 | 1 | .000 |
| Step 3 | Black | -2990.168 | 7.920 | 1 | .005 |
| | AIG | -2991.465 | 10.513 | 1 | .001 |
| | SWD | -2996.388 | 20.358 | 1 | .000 |
| Step 4 | Black | -2988.679 | 9.113 | 1 | .003 |
| | Hispanic | -2986.209 | 4.172 | 1 | .041 |
| | AIG | -2988.461 | 8.675 | 1 | .003 |
| | SWD | -2993.808 | 19.370 | 1 | .000 |
| Step 5 | Black | -2987.176 | 9.972 | 1 | .002 |
| | Hispanic | -2984.802 | 5.224 | 1 | .022 |
| | Mixed | -2984.123 | 3.867 | 1 | .049 |
| | AIG | -2986.254 | 8.129 | 1 | .004 |
| | SWD | -2991.900 | 19.422 | 1 | .000 |

a. Based on conditional parameter estimates

**Variables not in the Equation[a]**

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 1 | Variables | Asian | 1.727 | 1 | .189 |
| | | Black | 9.200 | 1 | .002 |
| | | Hispanic | 4.606 | 1 | .032 |

| | | | | | |
|---|---|---|---|---|---|
| | | Indian | .874 | 1 | .350 |
| | | Mixed | 2.561 | 1 | .110 |
| | | White | 10.141 | 1 | .001 |
| | | PacificIs | .396 | 1 | .529 |
| | | F | .008 | 1 | .928 |
| | | M | .008 | 1 | .928 |
| | | AIG | 11.447 | 1 | .001 |
| | | SES | 6.149 | 1 | .013 |
| | | ELL | 1.471 | 1 | .225 |
| Step 2 | Variables | Asian | 2.064 | 1 | .151 |
| | | Black | 8.086 | 1 | .004 |
| | | Hispanic | 2.996 | 1 | .083 |
| | | Indian | .928 | 1 | .335 |
| | | Mixed | 2.288 | 1 | .130 |
| | | White | 7.219 | 1 | .007 |
| | | PacificIs | .433 | 1 | .511 |
| | | F | .006 | 1 | .939 |
| | | M | .006 | 1 | .939 |
| | | SES | 3.679 | 1 | .055 |
| | | ELL | .729 | 1 | .393 |
| Step 3 | Variables | Asian | 1.644 | 1 | .200 |
| | | Hispanic | 4.200 | 1 | .040 |
| | | Indian | .956 | 1 | .328 |
| | | Mixed | 2.849 | 1 | .091 |
| | | White | 3.746 | 1 | .053 |
| | | PacificIs | .404 | 1 | .525 |
| | | F | .003 | 1 | .958 |
| | | M | .003 | 1 | .958 |
| | | SES | 3.237 | 1 | .072 |
| | | ELL | 1.135 | 1 | .287 |

| | | | | | |
|---|---|---|---|---|---|
| Step 4 | Variables | Asian | 1.047 | 1 | .306 |
| | | Indian | 1.012 | 1 | .315 |
| | | Mixed | 3.924 | 1 | .048 |
| | | White | .527 | 1 | .468 |
| | | PacificIs | .353 | 1 | .552 |
| | | F | .003 | 1 | .958 |
| | | M | .003 | 1 | .958 |
| | | SES | 1.770 | 1 | .183 |
| | | ELL | .050 | 1 | .824 |
| Step 5 | Variables | Asian | .723 | 1 | .395 |
| | | Indian | 1.050 | 1 | .305 |
| | | White | .673 | 1 | .412 |
| | | PacificIs | .323 | 1 | .570 |
| | | F | .004 | 1 | .950 |
| | | M | .004 | 1 | .950 |
| | | SES | 1.481 | 1 | .224 |
| | | ELL | .069 | 1 | .793 |

a. Residual Chi-Squares are not computed because of redundancies.

## APPENDIX H

**Binary Logistic Regression – Grades 4-8 Math**
**Dependent Variable Bonus Coded as 0: (MTHB _0)**

### Case Processing Summary

| Unweighted Cases[a] | | N | Percent |
|---|---|---|---|
| Selected Cases | Included in Analysis | 4490 | 100.0 |
| | Missing Cases | 1 | .0 |
| | Total | 4491 | 100.0 |
| Unselected Cases | | 0 | .0 |
| Total | | 4491 | 100.0 |

a. If weight is in effect, see classification table for the total number of cases.

### Dependent Variable Encoding

| Original Value | Internal Value |
|---|---|
| 0 | 0 |
| 1 | 1 |

## Block 0: Beginning Block

### Classification Table[a,b]

| | Observed | | Predicted MTHB_0 0 | Predicted MTHB_0 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 0 | MTHB_0 | 0 | 0 | 1752 | .0 |
| | | 1 | 0 | 2738 | 100.0 |
| | Overall Percentage | | | | 61.0 |

a. Constant is included in the model.
b. The cut value is .500

### Variables in the Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 0 | Constant | .446 | .031 | 212.964 | 1 | .000 | 1.563 |

## Variables not in the Equation[a]

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | Asian | 1.335 | 1 | .248 |
| | | Black | 7.762 | 1 | .005 |
| | | Hispanic | 5.130 | 1 | .024 |
| | | Indian | .964 | 1 | .326 |
| | | Mixed | 2.414 | 1 | .120 |
| | | White | 10.520 | 1 | .001 |
| | | PacificIs | .322 | 1 | .571 |
| | | F | .113 | 1 | .737 |
| | | M | .113 | 1 | .737 |
| | | AIG | 6.612 | 1 | .010 |
| | | SWD | 13.993 | 1 | .000 |
| | | SES | 4.754 | 1 | .029 |
| | | ELL | .642 | 1 | .423 |

a. Residual Chi-Squares are not computed because of redundancies.

## Block 1: Method = Forward Stepwise (Conditional)

## Omnibus Tests of Model Coefficients

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 14.249 | 1 | .000 |
| | Block | 14.249 | 1 | .000 |
| | Model | 14.249 | 1 | .000 |
| Step 2 | Step | 11.577 | 1 | .001 |
| | Block | 25.826 | 2 | .000 |
| | Model | 25.826 | 2 | .000 |
| Step 3 | Step | 7.919 | 1 | .005 |
| | Block | 33.745 | 3 | .000 |
| | Model | 33.745 | 3 | .000 |

| | | | | |
|---|---|---|---|---|
| Step 4 | Step | 4.171 | 1 | .041 |
| | Block | 37.916 | 4 | .000 |
| | Model | 37.916 | 4 | .000 |
| Step 5 | Step | 3.867 | 1 | .049 |
| | Block | 41.783 | 5 | .000 |
| | Model | 41.783 | 5 | .000 |

## Model Summary

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 5991.913[a] | .003 | .004 |
| 2 | 5980.337[a] | .006 | .008 |
| 3 | 5972.417[a] | .007 | .010 |
| 4 | 5968.246[a] | .008 | .011 |
| 5 | 5964.379[a] | .009 | .013 |

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

## Classification Table[a]

| Observed | | | Predicted MTHB 0 | Predicted MTHB 1 | Percentage Correct |
|---|---|---|---|---|---|
| Step 1 | MTHB | 0 | 0 | 1752 | .0 |
| | | 1 | 0 | 2738 | 100.0 |
| | Overall Percentage | | | | 61.0 |
| Step 2 | MTHB | 0 | 0 | 1752 | .0 |
| | | 1 | 0 | 2738 | 100.0 |
| | Overall Percentage | | | | 61.0 |
| Step 3 | MTHB | 0 | 68 | 1684 | 3.9 |
| | | 1 | 68 | 2670 | 97.5 |
| | Overall Percentage | | | | 61.0 |

| | | | | | |
|---|---|---|---|---|---|
| Step 4 | MTHB | 0 | 68 | 1684 | 3.9 |
| | | 1 | 68 | 2670 | 97.5 |
| | Overall Percentage | | | | 61.0 |
| Step 5 | MTHB | 0 | 68 | 1684 | 3.9 |
| | | 1 | 68 | 2670 | 97.5 |
| | Overall Percentage | | | | 61.0 |

a. The cut value is .500

## Variables in the Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | SWD | .320 | .086 | 13.918 | 1 | .000 | 1.376 |
| | Constant | .396 | .033 | 142.121 | 1 | .000 | 1.486 |
| Step 2[b] | AIG | .272 | .080 | 11.411 | 1 | .001 | 1.312 |
| | SWD | .377 | .087 | 18.670 | 1 | .000 | 1.458 |
| | Constant | .336 | .038 | 79.917 | 1 | .000 | 1.399 |
| Step 3[c] | Black | -.408 | .144 | 7.993 | 1 | .005 | .665 |
| | AIG | .259 | .081 | 10.368 | 1 | .001 | 1.296 |
| | SWD | .390 | .088 | 19.823 | 1 | .000 | 1.476 |
| | Constant | .356 | .038 | 86.346 | 1 | .000 | 1.427 |
| Step 4[d] | Black | -.440 | .145 | 9.205 | 1 | .002 | .644 |
| | Hispanic | -.166 | .081 | 4.195 | 1 | .041 | .847 |
| | AIG | .238 | .081 | 8.571 | 1 | .003 | 1.269 |
| | SWD | .381 | .088 | 18.877 | 1 | .000 | 1.464 |
| | Constant | .392 | .042 | 85.967 | 1 | .000 | 1.480 |
| Step 5[e] | Black | -.462 | .146 | 10.079 | 1 | .001 | .630 |
| | Hispanic | -.188 | .082 | 5.255 | 1 | .022 | .829 |
| | Mixed | -.248 | .125 | 3.908 | 1 | .048 | .781 |
| | AIG | .231 | .081 | 8.036 | 1 | .005 | 1.260 |
| | SWD | .382 | .088 | 18.927 | 1 | .000 | 1.465 |
| | Constant | .414 | .044 | 89.377 | 1 | .000 | 1.513 |

a. Variable(s) entered on step 1: SWD.

b. Variable(s) entered on step 2: AIG.

c. Variable(s) entered on step 3: Black.

d. Variable(s) entered on step 4: Hispanic.

e. Variable(s) entered on step 5: Mixed.

## Model if Term Removed[a]

| Variable | | Model Log Likelihood | Change in -2 Log Likelihood | df | Sig. of the Change |
|---|---|---|---|---|---|
| Step 1 | SWD | -3003.083 | 14.253 | 1 | .000 |
| Step 2 | AIG | -2995.958 | 11.579 | 1 | .001 |
| | SWD | -2999.746 | 19.155 | 1 | .000 |
| Step 3 | Black | -2990.168 | 7.920 | 1 | .005 |
| | AIG | -2991.465 | 10.513 | 1 | .001 |
| | SWD | -2996.388 | 20.358 | 1 | .000 |
| Step 4 | Black | -2988.679 | 9.113 | 1 | .003 |
| | Hispanic | -2986.209 | 4.172 | 1 | .041 |
| | AIG | -2988.461 | 8.675 | 1 | .003 |
| | SWD | -2993.808 | 19.370 | 1 | .000 |
| Step 5 | Black | -2987.176 | 9.972 | 1 | .002 |
| | Hispanic | -2984.802 | 5.224 | 1 | .022 |
| | Mixed | -2984.123 | 3.867 | 1 | .049 |
| | AIG | -2986.254 | 8.129 | 1 | .004 |
| | SWD | -2991.900 | 19.422 | 1 | .000 |

a. Based on conditional parameter estimates

## Variables not in the Equation[a]

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 1 | Variables | Asian | 1.727 | 1 | .189 |
| | | Black | 9.200 | 1 | .002 |
| | | Hispanic | 4.606 | 1 | .032 |

| | | | | | |
|---|---|---|---|---|---|
| | | Indian | .874 | 1 | .350 |
| | | Mixed | 2.561 | 1 | .110 |
| | | White | 10.141 | 1 | .001 |
| | | PacificIs | .396 | 1 | .529 |
| | | F | .008 | 1 | .928 |
| | | M | .008 | 1 | .928 |
| | | AIG | 11.447 | 1 | .001 |
| | | SES | 6.149 | 1 | .013 |
| | | ELL | 1.471 | 1 | .225 |
| Step 2 | Variables | Asian | 2.064 | 1 | .151 |
| | | Black | 8.086 | 1 | .004 |
| | | Hispanic | 2.996 | 1 | .083 |
| | | Indian | .928 | 1 | .335 |
| | | Mixed | 2.288 | 1 | .130 |
| | | White | 7.219 | 1 | .007 |
| | | PacificIs | .433 | 1 | .511 |
| | | F | .006 | 1 | .939 |
| | | M | .006 | 1 | .939 |
| | | SES | 3.679 | 1 | .055 |
| | | ELL | .729 | 1 | .393 |
| Step 3 | Variables | Asian | 1.644 | 1 | .200 |
| | | Hispanic | 4.200 | 1 | .040 |
| | | Indian | .956 | 1 | .328 |
| | | Mixed | 2.849 | 1 | .091 |
| | | White | 3.746 | 1 | .053 |
| | | PacificIs | .404 | 1 | .525 |
| | | F | .003 | 1 | .958 |
| | | M | .003 | 1 | .958 |
| | | SES | 3.237 | 1 | .072 |
| | | ELL | 1.135 | 1 | .287 |

| | | | | | |
|---|---|---|---|---|---|
| Step 4 | Variables | Asian | 1.047 | 1 | .306 |
| | | Indian | 1.012 | 1 | .315 |
| | | Mixed | 3.924 | 1 | .048 |
| | | White | .527 | 1 | .468 |
| | | PacificIs | .353 | 1 | .552 |
| | | F | .003 | 1 | .958 |
| | | M | .003 | 1 | .958 |
| | | SES | 1.770 | 1 | .183 |
| | | ELL | .050 | 1 | .824 |
| Step 5 | Variables | Asian | .723 | 1 | .395 |
| | | Indian | 1.050 | 1 | .305 |
| | | White | .673 | 1 | .412 |
| | | PacificIs | .323 | 1 | .570 |
| | | F | .004 | 1 | .950 |
| | | M | .004 | 1 | .950 |
| | | SES | 1.481 | 1 | .224 |
| | | ELL | .069 | 1 | .793 |

a. Residual Chi-Squares are not computed because of redundancies.

## APPENDIX I

**Letter of Agreement for the Use of Student Data**

June 1, 2019

To the Appalachian Institutional Review Board (IRB): I am familiar with Ross Rumbaugh's research project entitled: TEACHER MERIT PAY IN A RURAL WESTERN NORTH CAROLINA COUNTY: A QUANTITATIVE ANALYSIS OF THE EFFECTS OF STUDENT CHARACTERISTICS ON A TEACHER'S LIKELIHOOD OF RECEIVING A MONETARY BONUS IN MATH OR READING IN GRADES THREE - EIGHT

I understand that student data will be used for this research project.

Student data will be collected from the Secure Shell data base that houses data on tested students in North Carolina's public schools. The following data will be collected on all students in grades 3-8 that participated in end of grade math and reading testing in the 2017-2018 school year.

> ***Ethnicity*** – Asian students, African Americans, Hispanic, Mixed Race, Pacific Islander, White.
> ***Socio-economic level*** – The data available for this characteristic is dependent on the child's participation in the free and reduced lunch program.
> ***Sex*** – Male or Female (as indicated by parents on student data sheets)
> ***English as a second language program participant*** – Student participation in the English as a Second Language or EL program.
> ***Academically and intellectually gifted*** – Students participating in the AIG program.
> ***Exceptional children's program participants*** – Students served by the special education program.

The following will be collected from the school system's finance department:

> ***Teachers receiving the merit based bonus***– In North Carolina, the top 25% of teachers in grades four through eight received a bonus as a result of their students' growing academically in math. In grades three through five, the top 25% of teachers impacting student learning as measured by EVAAS also earned a bonus.

I understand that the informed consent of the student for the use of their data for the research project ***will not be*** obtained for this research but consent for gathering the student data will be received from the system superintendent.

As the research team conducts this research project the student data will be protected by:

- ***The researcher will not be involved with collecting the data. A testing and accountability assistant with the school system will gather the data and will mask all student and teacher names.***
- ***No data will be shared outside the research team including the testing assistant, Ross Rumbaugh (researcher), and Les Bolt (dissertation chair.) In the case of Mr. Rumbaugh and Dr. Bolt, they will not see student names or student ID's or teacher names as they will be removed before any statistical analysis is performed.***

- ***This data can be found in the state's Secure Shell file that can be accessed by certain public school officials and in PowerSchool which is available to teachers on a need-to-know basis. However, the data compiled for the research will be electronically and physically deleted after the analysis of the data is complete.***
- ***The data will be stored securely on a school system laptop and backed up on thumb drive. As previously noted, both electronic files will be deleted upon completion of the research with supervision of the superintendent of the county where the research is taking place.***
- ***No data will be shared outside the research team. With the student names, student ID numbers, and teacher names removed, no additional identifying information will be seen relating to the study participants' data.***
- ***The superintendent's name and school system will be masked in the dissertation so that the school system will remain anonymous.***

Therefore, as a representative of Appalachian State University, I agree that Ross Rumbaugh's research project may be conducted with the student data described above. I understand that this research will be carried out following sound ethical principles and that it has been approved by the IRB at Appalachian State University.

Sincerely,

Dr. Les Bolt, Dissertation Chair

Appalachian State University

## Vita

Ross Rumbaugh was born in Lancaster, Pennsylvania to Ross R. and Sharon L. Rumbaugh. He moved to North Carolina in 1988 to attend college at Wake Forest University. He graduated from Wake Forest University in May of 1992 with a Bachelor of Arts degree with honors in Economics. He received his teaching licensure from Winston-Salem State University in May of 1995. Additional degrees earned include a Masters of School Administration from Appalachian State University in May of 2003, and an Educational Specialist degree from ASU in May of 2012. In that same year he began the pursuit of his Doctorate in Educational Leadership from Appalachian State University where he will finish in December of 2019.

He has worked for 25 years in North Carolina's public school system in the roles of teacher, coach, assistant principal, principal, Director of Administrative Services, and Assistant Superintendent. He currently serves as the Director of Testing and Accountability for Burke County Public Schools in Morganton, North Carolina.

Dr. Rumbaugh currently resides with his wife Marissa of 21 years and his children Ellie and Brittain Rumbaugh of whom he is most proud. He serves as a deacon at Corinth Reformed Church in Hickory, North Carolina. He also coaches basketball and lacrosse and provides basketball clinics for local youth coaches.