Spring 3-19-2019

The Predictive Power of Teacher Practice in Explaining Student Growth

Nick Pillsbury
nicholas.pillsbury@student.shu.edu

The Predictive Power of Teacher Practice in Explaining Student Growth

by

Nick Pillsbury

Submitted in Partial Fulfillment of the Requirements for the degree of

Doctor of Education

Department of Education, Management, Leadership and Policy

Seton Hall University

May 2019

SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATE STUDIES

APPROVAL FOR SUCCESSFUL DEFENSE

Nick Pillsbury has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Spring Semester 2019**.

DISSERTATION COMMITTEE

(please sign and date beside your name)

Mentor:
Dr. Elaine Walker Elaine M-Walker 3/19/19

Committee Member:
Dr. Martin Finkelstein 3/19/19

Committee Member:
Dr. Peter Mazzagatti Peter Mazzagatti 3.19.19

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file and submit a copy with your final dissertation to be bound as page number two.

ABSTRACT

In recent years, states across the nation have increased their interest in developing specific teacher accountability measures and improving student achievement. On August 6, 2012, the state of New Jersey approved the TEACHNJ Act, which reformed tenure laws and linked student growth to a teacher's evaluation. The ultimate goal of the TEACHNJ Act is to "raise student achievement by improving instruction through the adoption of evaluations that provide specific feedback to educators" (TEACHNJ Guide, 2012). The 2013-14 school year was the first full year of implementation and included student growth percentile (SGP) scores as one component of a teacher's evaluation.

The purpose of this quantitative study was to examine the relationship between teacher practice and student growth. The study determined the probability that a student will have typical or high growth on the state assessment in relation to the teacher's practice score based on classroom observations. Some of the essential questions regarding this research are as follows: a) Are teacher-level variables such as gender, ethnic background, and age significant predictors of student growth? b) Are school-level variables such as school performance status (Comprehensive schools, Target schools, and NonStatus schools) and percent of student subgroup ethnic composition significant predictors of student growth? c) How is student growth in language arts and mathematics impacted by a teacher's effectiveness as the practice score measures it when one controls for teacher- and school-level characteristics? and d) To what extent does the relationship between teacher effectiveness and student outcome vary from year 2 of AchieveNJ to year 5 of AchieveNJ?

The sample population for the 2014-2015 school year will consist of 1,132 students (n = 1,132) with a valid language arts SGP and 1,087 students (n = 1087) with a valid mathematics

SGP. The sample population for the 2017-2018 school year consisted of 1,484 students (n = 1,484) with a valid language arts SGP and 1,473 students (n = 1,473) with a valid mathematics SGP. The study involved 12 to 14 schools with different grade configurations, performance status, and student ethnic composition.

This study was a cross-sectional explanatory design in which logistic and hierarchical logistic regression methods were used to test the relationships between the dependent variable (student growth) and independent variables (teacher characteristics, school characteristics, and teacher practice). The design consisted of three separate models used to answer four research questions. A logistic regression analysis will be used to analyze Model 1 (teacher characteristics on student growth) and Model 2 (school characteristics on student growth). In Model 3, a hierarchical logistic regression analysis was used to better interpret the impact of teacher practice and teacher and school characteristics on student growth. Research question four compared and analyzed the significant findings between Model 3 in the 2014-2015 school year and Model 3 in the 2017-2018 school year.

This study will provide insight for educational leaders and policymakers on the positive relationship between teacher practice and student growth. It also recommends that this type of research continue to explore how other variables influence student learning growth based on how teachers deliver instruction.

DEDICATION

Most importantly, this work is dedicated to my wife, Ashley Pillsbury, who has sacrificed an exorbitant amount of her time throughout this process. It is through your encouragement, support, and selflessness that I was able to complete this endeavor.

Second, I dedicate this work to my children, Grayson and Sutton, who were an inspiration to me during this experience and were one of the main reasons I took on this venture to instill in them that education is a powerful tool, and never to give up on a goal.

ACKNOWLEDGEMENTS

I especially acknowledge my mentor, Dr. Elaine Walker, for your knowledge, guidance, and commitment to my work throughout this lengthy process. You encouraged and motivated me to complete this work, and for that I am truly grateful. A special thanks to my committee members, Dr. Martin Finkelstein and Dr. Peter Mazzagatti, for your knowledge, contribution, and interest in my work.

I would like to acknowledge my Seton Hall Ed.S. – Ed.D. Cohort III members who began this process with me, especially Kevin Jarido, who started at the very beginning with me in the Seton Hall Education Leadership, Management, and Policy Camden Cohort II. Thank you all for your support, knowledge, and enjoyment throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Policy Background

Teacher evaluation is not a new concept. Even over 30 years ago, a great deal of national attention focused on setting well-designed evaluation processes to identify and promote highly effective teaching. In 1983, when the National Commission on Excellence in Education published *A Nation at Risk: The Imperative for Educational Reform*, it created a new interest in teacher evaluation processes. The commission recommended, "Persons preparing to teach should be required to meet high educational standards, to demonstrate an aptitude for teaching, and to demonstrate competence in an academic discipline...Salary, promotion, tenure, and retention decisions should be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated" (Wise, 1984, pg. 18).

Wise (1984) broke down the concept of teacher evaluations into four basic purposes: individual staff development, school improvement, individual personnel decisions, and school status decisions. The first two purposes take a formative approach and involve improving practice; the second two purposes take a summative approach and focus on accountability to measure teacher effectiveness. Although many teacher evaluation systems seek to accomplish all four of these purposes, different processes and methods may better suit individual objectives. In particular, an evaluation system that focuses on improvement or accountability requires different standards of acceptability.

Wise (1984) explained that evaluation systems that focus on accountability must be capable of yielding standardized and objective evidence about a teacher's performance. For

teacher evaluation systems that focus on improvement, more emphasis must be placed on descriptive information that identifies areas for growth and produces a course for change. According to the researcher, the improvement of a teacher relies mostly on the success of two conditions. One is for both the teacher and supervisor to have crafted a course of action that they both agree is the correct one to produce improved practice. The other condition necessary for teachers to improve their practice is for them to feel empowered and confident that the given course is possible to achieve.

Although the importance of teacher effectiveness has been discussed and debated for many years, it was in the turn of the 21st century, with the No Child Left Behind Act (NCLB), when the federal government stepped up and applied a significant amount of pressure on states to improve student proficiency and close the achievement gap. NCLB is the most recent update to the Elementary and Secondary Education Act (ESEA) of 1965, which was the birth of Title I and provided supplemental funding to school districts with a greater number of disadvantaged students (Editorial Project in Education Research Center, 2015). NCLB held schools responsible for the academic progress of each student. By the 2013-2014 school year, every student had to reach a "proficiency level" that the state set. In order to track progress toward the end goal, schools were required to make "adequate yearly progress," or AYP. There were several consequences if a school did not meet that mark every year, which included allowing students to transfer to better performing schools in the same district, offering free tutoring, and facing state intervention. Schools that failed to achieve the required level of academic proficiency would also risk losing their Title I funds (Editorial Project in Education Research Center, 2015). All of these factors put a significant amount of pressure on school districts with the lowest levels of proficiency.

By 2010, 38% of schools were not meeting AYP, and it was clear that many of these schools would not be able to reach the achievement targets that the NCLB set (Editorial Project in Education Research Center, 2015). During that year the Obama administration offered several waivers to schools that were not going to meet NCLB target. However, states with schools that were awarded this waiver "had to agree to set standards aimed at preparing students for higher education and the workforce" (p. 5). Some states chose to adopt the Common Core State Standards, but others set standards that higher education institutions within their state approved. Another requirement was for these states to establish and implement teacher evaluation systems that used student progress on standardized tests as a measurement for teacher effectiveness.

One year prior, in 2009, President Obama had signed the American Recovery and Reinvestment Act (ARRA), which was designed to support job creation and invest in critical sectors such as the education department (U.S. Department of Education, 2009). From this act, the Race to the Top Fund was born. Race to the Top is a grant program that awarded states for creating innovative education reform that could demonstrate the ability to make significant growth in student outcomes by closing the achievement gap. For a state to be awarded this grant, they had to have met six general criteria. The criterion that was emphasized the most, and awarded the most points, was "Great Teacher and Leaders." States must show how they would improve teacher and principal effectiveness based on performance.

The state of New Jersey was one of the states awarded this grant and began the work to improve educator evaluations to increase teacher and leader effectiveness. To create an effective evaluation system, the New Jersey Department of Education established an Educator Effectiveness Task Force to identify what an effective evaluation system should look like. In 2011, the Task Force released a report that outlined steps for executing an improved evaluation

system for teachers and principals. One of those steps was to pilot the new evaluation procedures for 2 years and include more than 15,000 teachers and principals (New Jersey Educator Effectiveness Task Force, 2011).

In 2012, New Jersey's State Legislature unanimously approved the TEACHNJ Act, which set mandates for the new educator evaluation system and made significant changes to tenure decisions. Teacher and principals would now have their tenure determined by their evaluation rating.

The law declared:

> The New Jersey Supreme Court has found that a multitude of factors play a vital role in the quality of a child's education, including effectiveness in teaching methods and evaluations. Changing the current evaluation system to focus on improved student outcomes, including objective measures of student growth, is critical to improving teacher effectiveness, raising student achievement, and meeting the objectives of the federal "No Child Left Behind Act of 2001" (Teacher Effectiveness and Accountability for the Children of New Jersey [TEACHNJ] Act, Chapter 26, 2, 2012, p. 1).

The ultimate goal of the TEACHNJ Act is to "raise student achievement by improving instruction through the adoption of evaluations that provide specific feedback to educators, inform the provision of aligned professional development, and inform personnel decisions" (Teacher Effectiveness and Accountability for the Children of New Jersey (TEACHNJ) Act, Chapter 26, 2, 2012, p. 1). New Jersey implemented its 1st year of the new evaluation policies for the 2013-14 school year after the State Board of Education approved the specific regulations.

These new evaluation policies were a major shift from prior evaluation procedures. The policies eliminated the binary method of evaluation (i.e., satisfactory vs. unsatisfactory), which did not take into account student growth. The new system, however, now incorporates student growth along with a teacher's practice to produce an overall performance score that reflects one of four different teacher rating categories (highly effective, effective, partially effective, and ineffective). Prior to the new law, tenured teachers might have only received one observation for the entire year. TEACHNJ requires multiple observations for all teachers. Table 1 provides an overview of the evaluation process prior to and after the implementation of the new law.

Table 1

New Jersey Teacher Evaluation Framework

Teacher Evaluation Prior to AchieveNJ (Past)	**Teacher Evaluation–AchieveNJ (Present)**
Binary measurement with limited ability to differentiate effectiveness and inform growth	Four-tiered measurement to differentiate levels of effectiveness and inform growth
Evaluation based solely on single measure (teacher practice)	Evaluation based on multiple measures (teacher practice and student achievement)
Multiple observations (3) required for non-tenured teachers	Multiple observations required for all teachers
	Ongoing calibration and monitoring of observations

(Adapted from "Teacher Effectiveness and Accountability for the Children of New Jersey (TEACHNJ) Guide," 2014, p. 3).

Before the TEACHNJ Act, districts had the discretion to use their own methods to assess teacher practice, which might be through a district-designed practice rubric or as simple as a checklist. Currently, districts must select from a list of state-approved teacher practice evaluation instruments in accordance with the TEACHNJ Act. The teacher practice evaluation instrument assesses the competencies of a teacher's practice by gathering evidence, primarily through classroom observations. These state-approved instruments include Charlotte Danielson

Framework for Teaching, Marzano's Causal Teacher Evaluation Model, Mid-Continent Research for Education and Learning (McREL) Teacher Evaluation Standards, and Focal Point Teaching Practice Model. School districts can change or revise their selected teacher practice evaluation instrument each year but must follow specific state guidelines to do so. At the time of this study, AchieveNJ requires tenured teachers to be observed two times annually. Nontenured teachers are to be observed three times annually. In addition, more than one certified administrator (multiple observers) are required to observe nontenured teachers and teachers placed on a corrective action plan. Teachers are automatically placed on a corrective action plan when they are rated ineffective or partially effective on their summative evaluation. As per New Jersey state statute:

> For each teaching staff member rated ineffective or partially effective on the annual summative evaluation, as measured by the evaluation rubrics, a corrective action plan shall be developed by the teaching staff member and the teaching staff member's designated supervisor. If the teaching staff member does not agree with the corrective action plan's content, the designated supervisor shall make the final determination (Educators Effectiveness, 2012, p. 25).

In addition to the teacher's practice evaluation score, student achievement measures are calculated and incorporated into a teacher's summative evaluation. Student achievement measures consist of student growth percentiles (SGPs) and/or student growth objectives (SGOs). SGPs measure student achievement gains in grades 4–8 in language arts and in grades 4–7 in mathematics (tested grades and subjects) on the Partnership for Assessment of Readiness for College and Careers (PARCC) state assessment, which was adopted and administered to all

students in New Jersey in the spring of 2014. An SGP score is a number on a scale from 1 to 99 that measures the change in a student's achievement from one year to the next compared to all other students, or "academic peers," in the state who had similar historical results (New Jersey Department of Education, 2015). A student's SGP growth can be categorized as low (SGP < 35), typical (SGP > 34 and SGP < 66), or high (SGP > 65; NJ SMART Education Data System, 2016). Teacher's evaluations reflect the median student growth percentile (mSGP) of all the students in their class. Students below grade level with a low proficiency rate could earn a high SGP score, which means that these students demonstrated more growth than their "academic peers" and signals that the teacher's instructional practice may have assisted the students' growth.

During both the 2014–2015 and the 2017–2018 school years, a teacher's overall summative evaluation rating includes the following multiple measures: (a) teacher practice score derived from two or three observations, and (b) SGP scores and/or SGO scores, which are weighted and added together to calculate an overall summative evaluation score from 1 (ineffective) to 4 (highly effective). Teachers in tested grades and subjects receive an mSGP score, and the overall summative evaluation rating in the 2014-2015 school year was calculated by combining the multiple weighted measures of teacher practice (70%), mSGP (10%), and SGO (20%). For teachers in nontested grades and subjects during the 2014-2015 school year who did not receive an mSGP score, the overall summative evaluation rating was calculated by combining teacher practice (80%) and the average of two SGOs (20%). During the 2017-2018 school year, on the other hand, the overall summative evaluation rating for teachers in tested grades and subjects was calculated by combining the multiple weighted measures of teacher practice (55%), mSGP (30%), and SGO (15%). For teachers in nontested grades and subjects

during the 2017-2018 school year who did not receive an mSGP score, the overall summative evaluation rating was calculated by combining teacher practice (85%) and the average of two SGOs (15%).

The district adopted the Department of Education's state-approved Charlotte Danielson Framework for Teaching instrument to evaluate teacher practices. The instrument focuses on the following four performance domains:

- Planning and Preparation
- Classroom Environment
- Instruction
- Professional Responsibilities

Statement of the Problem

The new teacher evaluation requirements have forced teachers to adapt to a new model of accountability, which determines teacher effectiveness by establishing a relationship between student achievement and teacher evaluations. Currently, no substantial amount of research in the state has offered insight to examine the link between teacher practice and SGPs. However, there have been several studies that illustrated inconsistencies between an individual teacher's rating and student performance. Weisberg, Sexton, Mulhern, and Keeling (2009) discovered that many educational policies in the United States carry the assumption that teachers are interchangeable. With this assumption, "school systems wrongly conflate educational access with educational *quality*" (Weisberg et al., 2009, p. 9). This phenomenon is what is known as the Widget Effect. The study goes on to identify how the implementations of many new evaluation systems continue to reinforce the Widget Effect.

Other research has illustrated how extreme fluctuations in teacher evaluation rating from year to year have raised validity issues with the reliability of new teacher evaluation systems. Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein's (2012) study examined teacher evaluation data from five school districts and found that only 20% to 30% of teachers who were rated below the effective rating in the 1st year were rated the same the following year, compared to 25% to 45% of teachers who were rated below effective but moved all the way to the highly effective rating the following year. In contrast, only a small minority of highly effective teachers remained highly effective the following year. This study will continue to evaluate the impact of the TEACHNJ Act and examine the relationship between teacher practice and student growth in another school district that the Department of Education has identified as being faithful in its implementation of the new evaluation system.

Purpose of the Study

In 2012, the state of New Jersey mandated the implementation of its new teacher evaluation system. During the 1st full year of implementation of the TEACHNJ Act, a study was conducted to identify the relationship between teacher practice and student growth. One of the recommendations for future research is for a similar study to be replicated in another state-controlled district to determine if the results are similar. This study will be completed during the 3rd full year of implementation of the new teacher evaluation mandates.

The purpose of this study was to examine the relationship between teacher practice and student growth while controlling for teacher- and school-level characteristics. The study determined the probability that students will receive a typical or high growth SGP score on the state assessment in relation to their teacher's practice score based on the practice portion of the evaluation instrument. The study will determine the value added by teacher practice, teacher

characteristics, and school characteristics on student achievement in the content areas of language arts and mathematics in grades 4–7. The state reports the change in student growth as an SGP score on a scale from 1 to 99.

The first two research questions below were used to analyze the impact of teacher and school characteristics on student growth separately in Model 1 and Model 2; the third research question included the main predictor variable (teacher practice) in a hierarchical logistic regression in Model 3 to identify its value added on student growth.

Research Questions

1. Are teacher-level variables such as gender, ethnic background, and age significant predictors of student growth?
2. Are school-level variables such as school performance status (Comprehensive schools, Target schools, and NonStatus schools) and percent of student subgroup ethnic composition significant predictors of student growth?
3. How is student growth in language arts and mathematics impacted by a teacher's effectiveness as measured by the practice score received when one controls for teacher- and school-level characteristics?
4. To what extent does the relationship between teacher effectiveness and student outcome vary from year 2 of AchieveNJ to year 5 of AchieveNJ?

Significance of the Study

When the federal government initiated the Race to the Top grant program, New Jersey was one of the states awarded this grant to create innovative education reform, which includes the construction of a new teacher evaluation system that could identify and improve a teacher's effectiveness and ability to increase student achievement. A teacher's evaluation would now

incorporate multiple evaluation measures, provide specific feedback for improvement, be aligned to professional development, and inform tenure decisions.

At the time of this research study, the state of New Jersey was in its 4th year of implementation of the new teacher evaluation framework (AchieveNJ), and any research finding will benefit future policy decisions. This study will continue the previous research that examined the relationship between SGPs and teacher practice in a different school district in an urban setting with a large number of poor- and low-performing students.

Although the federal government and state administrations remain focused on student achievement on state assessments and teacher evaluations as measures of teacher effectiveness, the public policy debate will strengthen around the validity of using value-added measures for tenure, retention, and termination. More specifically, debates will focus on whether teachers should be evaluated based on student achievement, especially when some researchers are skeptical about using value-added measures. The skepticism is bolstered by the fact that it has been shown that teacher performance can fluctuate over time depending on several factors that influence student growth such as attendance rate, mobility rate, curriculum material, instructional time, class size, and home and community supports (Darling-Hammond et al., 2012).

Using the scores of students with like scores across the state of New Jersey to measure student academic growth makes this study unique. Students are compared to their "academic peers" to determine growth regardless of their level of proficiency, their socioeconomic background, and whether student assignments to teachers are randomized. Further, the study will explain the relationship between teacher practice and student growth in an urban school district that the state of New Jersey identified as a district in need of improvement.

Theoretical Framework

New teacher evaluation systems in many school districts signify an exit from the previous approach that typically used checklists with little observational feedback and rarely included data on student achievement (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). A study that Weisberg et al. conducted (2009) found that in 12 school districts across four states, less than 1% of teachers were rated unsatisfactory, teachers did not receive specific feedback on improving their practice, novice teachers were neglected, and poor performance went unaddressed. In reaction to the new evaluation mandates, controversy has surrounded the validity of using student achievement to assess the effectiveness of teacher practice (Kane, Taylor, Tyler, & Wooten, 2011). Should teachers be held accountable for student achievement on standardized tests when there may be other factors outside of their control?

The development of these new policies depends on the logic linking teacher evaluation to student achievement and has been based on several assumptions. Some studies have confirmed a direct relationship between teacher effectiveness and student academic success. Kane, McCaffrey, Miller, and Staiger (2013) designed the Measures of Effective Teaching (MET) project, which was essentially established to identify effective teachers. The MET project began by collecting baseline data on teacher effectiveness to predict the impact that a teacher will have on a different group of students the following year. To determine teacher effectiveness, the MET project combined three approaches that could measure an aspect of teaching (student survey, classroom observations, and a teacher's track record of student achievement gains on a state test). Once baseline data were collected, students were randomly assigned to teachers the following school year, and the researchers collected student achievement data to determine the relationship of student achievement and teacher effectiveness. The MET project compared

predicted student outcomes to the actual differences that emerged at the end of the 2nd school year. The study was able to determine that teachers who were identified as effective produced greater student growth as compared with teachers who were rated less than effective (Kane et al., 2013). Similar studies may have justified the use of student achievement data in teachers' performance evaluations. This data may also produce reliable and sustainable improvements in the quality of teaching and learning (Darling-Hammond et al., 2012).

To understand the influence of teacher practice on student growth, it must be assumed that several factors are likely to impact student performance. These assumptions, derived from the existing literature, justify that student growth is a function of three sets of variables: teacher practice (the main independent variable), teacher characteristics, and school characteristics. In addition, it can be argued that both teacher- and school-level characteristics also influence teacher practice.

Limitations of the Study

1. The study will exclude bilingual teachers because there were only a few valid teacher practice scores.
2. The teacher evaluation framework does not have an interrater reliability component, and there will be variance in professional development given to observers who provide feedback to teachers.

Delimitations of the Study

1. This study will focus on students who are in grades 4–7 in language arts or mathematics and receive a valid SGP score, which the state calculates and provides.
2. Data collection is confined to state assessment results and teacher practice scores from the school district's Teachscape and TeachBoost systems, which maintain teachers'

observation and evaluation scores. School-level data that include teacher and school characteristics are collected online from New Jersey State school performance reports, and data will be retrieved from the state's NJSMART portal, which maintains student and teacher records.

Definition of Terms

Academic Peers: students from around the state of New Jersey with similar score histories on state assessments.

AchieveNJ: a state mandate that relies on multiple measures of performance to evaluate teachers. These measures include components of both student achievement and teacher practice.

Evaluation Instrument: a teaching practice evaluation instrument that a school district selects from a state-approved list. The evaluation instrument is a rubric that provides measurements that capture teacher competencies.

Observation: a method of collecting data on the performance of a teaching staff member's responsibilities and that would be included in the determination of the annual summative evaluation rating.

Student Growth Objectives (SGOs): long-term academic goals for groups of students that teachers set in consultation with their supervisors.

Student Growth Percentiles (SGPs): New Jersey measures growth for an individual student by comparing a student's growth to the growth that the student's academic peers made within a testing year.

Summative Evaluation: consists of two primary components: teacher practice (measured primarily by classroom observations) and student achievement. Under AchieveNJ, teachers are

evaluated based on multiple measures of educator practice and student achievement. Each element of the evaluation results in a rating of 1 to 4, which is weighted according to the state formulas. Once the scores for all evaluation measures are finalized, each educator receives a final summative rating on a scale from 1 to 4 (1 = ineffective, 2 = partially effective, 3 = effective, 4 = highly effective).

Teacher Practice: the methods by which a classroom teacher delivers instruction.

TEACHNJ Act: the tenure reform law, which reformed the processes of earning and maintaining tenure. Under the act, tenure decisions are based on multiple measures of student achievement and teacher practice that the new evaluation procedures measure. All teachers have to earn an evaluation rating of effective or highly effective to maintain tenure. Any tenured teaching staff member who was rated ineffective or partially effective in two consecutive summative annual evaluations could be charged with inefficiency.

Value-Added Modeling (VAM): a method of teacher evaluation that measures the teachers' contribution in a given year by comparing the current test scores of their students to the scores of those same students in previous school years, as well as to the scores of other students in the same grade.

CHAPTER II

LITERATURE REVIEW

Overview

The literature review begins by examining how the nation has increased its focus on teacher accountability over the past few decades, and how states have implemented policies to include student growth measures as a means to evaluate teacher performance. The chapter also reviews the concept of teacher effectiveness and its relationship to student achievement. The literature review explores the new incentive to incorporate value-added models (VAMs) to determine teacher effectiveness and hold teachers accountable for student learning. These accountability measures require teacher evaluations to support effective teaching practices, to support the retention of effective teachers, and to encourage the dismissal of ineffective teachers. Review of the literature revealed that although some research findings have promoted the use of VAMs to determine teacher effectiveness for evaluation purposes, other research has identified setbacks in using VAMs alone to evaluate teachers' effectiveness. Much of the literature was empirical in nature rather than theoretical. This chapter takes a closer look at Student Growth Percentiles (SGPs), which New Jersey uses to measure student growth on its state summative assessment. The current climate of educational reform and current accountability processes that have been put into place in many school districts require a critical review of their impact on the education process.

Teacher Effectiveness and Student Achievement

As previously noted in Chapter 1, the Race to the Top Fund was a competitive grant program designed to "encourage and reward States that are creating the conditions for education innovation and reform; achieving significant improvement in student outcomes, including

making substantial gains in student achievement, closing achievement gaps, improving high school graduation rates, and ensuring student preparation for success in college and careers" (U.S. Department of Education, 2009, p. 2). Among the criteria that states had to meet to be awarded the Race to the Top grant, the criteria that was emphasized the most and awarded the most points was "Great Teacher and Leaders." For states to meet these criteria, they must show how they will improve teacher and principal effectiveness based on performance.

Among the many states that were awarded funds through the Race to the Top grant program was Illinois. In accordance with the Performance Evaluation Reform Act (PERA) of 2010, every district in Illinois would have to adopt a new teacher evaluation system that addressed both teacher performance and student growth. The teacher performance measures that PERA required had to include both formal and informal classroom observations. Various qualifying assessment types had to have defined student growth measures. Both teacher performance and student growth ratings had to be combined to create a single summative rating of teacher performance. PERA was phased in over the following 4 years, starting with the lowest performing schools, and progressing to statewide adoption by the beginning of the 2016-2017 school year (White et al., 2012).

Although student growth measures were now required to be connected to a teacher's evaluation, Sartain, Stoelinga, and Krone (2010) explained that achievement data should not be the only indicator of teachers' need to improve their practice. The federal government recognized that the other major component of teacher evaluations must include some form of classroom observation, which created a new demand for observation frameworks that principals and others could use to identify effective teaching. The Consortium on Chicago School Research (CCSR) conducted a study on the implementation of the Charlotte Danielson Framework for Teaching in

Chicago Public Schools (CPS), which was their observational tool used for formal teacher observations. This framework was a major shift from their previous observation tool, which was simply a "checklist" used for the previous 30 years and provided very little insight into effective teaching practice. The Danielson Framework, on the other hand, was able to delineate the observable components of effective teaching.

Cincinnati's Teacher Evaluation System (TES) conducted another study that evaluated the effectiveness of the Danielson Framework to demonstrate a relationship to student achievement. Kane et al. (2011) determined that classroom observations can capture elements of teaching that are related to student achievement. Their findings demonstrated a positive and nontrivial relationship between TES scores and student achievement growth. Their main results indicated that moving from an overall TES rating of "Basic" to "Proficient" or from "Proficient" to "Distinguished" is associated with student achievement gains of about one sixth to one fifth of a standard deviation. For example, if a student started the year at the 50th percentile in math and reading and had a teacher in the lowest quartile of the overall TES rating, while a similar student had a teacher in the upper quartile of that rating, we would expect the second student to be four percentile points ahead in math and five percentile points ahead in reading by the end of the year.

The first state, however, to create a value-added system to measure student growth was the Tennessee Department of Education. The Tennessee Value-Added Assessment System (TVAAS) could measure individual teacher performance to student test score gains. In 1985, a scientifically controlled experiment called Project STAR, which stood for Student/Teacher Achievement Ratio, was conducted to test the impact of class size on student achievement and, in particular, the achievement gains from smaller classes by race. Although it was determined that all students performed better in smaller class sizes, the study found that the gains in effect

size for minorities were approximately twice the gains of whites, which essentially reduced the achievement gap. The study randomly assigned more than 6,000 students from various racial and socioeconomic backgrounds to small (13–17 students) and large (22–26 students) classes in 79 schools across the state and offered an opportunity to examine differences in student achievement where the only difference between the classes was the teacher. In addition, teachers were randomly assigned to classrooms each year. Students were assigned to the same class size for up to 4 years. There were no interventions, no special training for teachers, and no special curricula. Achievement gains were greater each year for smaller classes than larger classes. The effect of small classes in mathematics for 3 years (grade 1 = 0.140, grade 2 = 0.063, and grade 3 = 0.067) yielded an average effect of 0.090 deviations per year. In reading over 3 years (grade 1 = 0.124, grade 2 = 0.076, and grade 3 = 0.112), small class size yielded an average of 0.104 standard deviations per year. The difference between small classes and large classes was 0.2 to 0.3 standard deviations in each subject. The study showed that the benefits for small classes were two to three times greater for minority students who attended inner city schools than for White students who attended suburban schools. In large classes, the achievement gap between White and Black students in reading was 14.3% compared to 4.1% in small-class settings (Finn, 2002). Finn indicated that at the end of grade 6, students who attended small classes for 1 year had a 1.2-month advantage in reading over students who attended large classes. Students who attended small classes for 2 years had a 2.8-month advantage, and those who attended for 3 years had a 4.4-month reading advantage over students who attended large-class settings.

Stronge, Ward, and Grant (2011) examined the characteristics of effective teachers versus less effective teachers by examining classroom instructional and management practices. Student learning gains were measured for 1 year where the grade 4 end-of-course reading and

mathematics tests served as the grade 5 pretests. The study comprised 1,984 students, of which 931 students were assigned to less effective teachers and 1,053 to effective teachers. The results of the study indicated that student achievement in language arts and mathematics was higher for effective teachers than for less effective teachers by more than 30 percentile points. For reading and mathematics, the difference in gains in 1 year was 0.59 and 0.45 standard deviations, respectively. Stronge et al. (2011) noted, "This translated into more than a 30 percentile difference in achievement based on 1 year's teaching and learning experience" (p. 345). The comparison of teacher practice between effective and less effective teachers did not reveal a significant difference in teacher beliefs, teacher questioning, student questioning, or student disengagement. However, the results indicated that less time was spent on task due to disruptive behaviors, and less effective teachers had three times more disruptions than effective teachers. In addition, data from observation ratings on teachers' effectiveness indicated statistically significant differences favoring the effective teachers on classroom management ($p<.01$), classroom organization ($p<.02$), positive relationships with their students (.03), and encouragement of student responsibility ($p<.01$).

These studies have shown that teacher effects on student learning as inferred from standardized test scores are additive and cumulative over grade levels, and that teacher effectiveness can be measured fairly.

Value-Added Models

As states move forward with establishing teacher evaluation systems, VAMs have become a key component for most of these systems to measure accurate teacher effectiveness. VAMs attempt to predict the "value" that teachers add to their students' learning growth as the standardized assessments measure. Some states have mandated that up to 50% of the teacher

evaluation be tied to student test scores using a value-added measure. The logic of using teacher evaluation to measure teacher effectiveness for school improvement is based on the positive relationship between teacher quality and student academic growth. Administrators collect data on teacher classroom behavior through classroom observations and compare the results against teacher practice standards on an identified teacher evaluation rubric. Student growth models are then used to measure the amount of growth a student shows from the previous year. This information determines retention, promotion, compensation, and tenure. The use of VAMs for these high-stakes consequential decisions has many questioning its reliability, validity, and consistency.

Hallinger, Heck, and Murphy (2014) conducted a critical evaluation of the empirical literature and found few studies that indicated benefits in using VAMs. A study that Taylor and Tyler conducted (2012) in a Cincinnati school district found evidence to suggests that midcareer teachers' effectiveness improved during the school year and subsequent following school years when VAMs were incorporated into the evaluation process. Students in mathematics performed higher on end-of-year math tests the year that value-added measures were in place compared to the previous year's evaluations. Taylor and Tyler explained:

> These improvements persist and, in fact, increase in the years after evaluation. We estimate that the average teacher's students score will be 0.11 standard deviations higher in years after the teacher has undergone an evaluation compared to how her students scored in the years before her evaluation. To get a sense of the magnitude of this impact, consider two students taught by the same teacher in different years, who both begin the year at the 50th percentile of math achievement. The student taught after the teacher went through the TES process

> would score about 4.5 percentile points higher at the end of the year than the student taught before the teacher went through the evaluation (p. 83).

Milanowski (2004) examined the teacher evaluation system in Cincinnati to determine the relationship between the evaluation scores of teachers and VAMs of student learning in grades 3–8. The school system's administrators "want[ed] to be justified in inferring that teachers with high scores [were] better performers, defined as producing more student learning" (p. 39). The study yielded some positive and mixed results. However, Milanowski determined that the "moderate level of criterion-related validity" (p. 49) was adequate to support the use of student achievement data in the evaluation of teachers.

Borman and Kimball (2005) studied a sample of 400 teachers and 7,000 students in a school district in Reno, NV. Their goal was to assess whether the standards-based evaluation system helped close the achievement gap among students of different socioeconomic backgrounds. Their results showed a higher mean achievement in classrooms that effective teachers taught, but the differences were not significant. The authors concluded:

> This analysis suggests that teacher quality, as defined and applied in the evaluation system of one school district, may not show reliable relations to closing achievement gaps between poor and more advantaged, minority and nonminority, and low- and high-achieving students. The implications for the evaluation system are important, especially if a key component of teacher quality is an ability to close achievement gaps (Borman & Kimball, 2005, p. 18).

The greatest variability in student outcomes can be attributed to the student's background and factors outside the control of teachers.

Kimball, White, Milanowski, and Borman (2004) conducted a larger scale study of a teacher evaluation system in Washoe County, Nevada in which they wanted to understand if "teachers who score well on such evaluation systems also help produce higher levels of student learning" (p. 56). This research examined the relationship between teacher evaluation results and student gains in achievement in reading and math. The results were mixed. The relationship between teacher evaluations scores to student achievement was positive in each grade in language arts and mathematics but was not statistically significant.

Additional studies found little significance when they examined the relationship between student achievement and teacher evaluation ratings. White's (2004) study in Coventry, Rhode Island sought to "describe the relationship between a teacher's overall evaluation score and his or her students' achievement, while controlling for prior achievement, in order to determine the criterion-related validity of the evaluation scores" (p. 3). The researcher analyzed the value-added achievement data in reading and math from 3,617 students and evaluation data for 173 teachers in four elementary school grades and for 2 school years. White's results "indicated a small overall correlation in reading (0.240) and essentially no correlation in math (0.032). The results also indicated rather large fluctuations in correlations between years and across subjects and grade levels" (p. 6). Again, the overall pattern of results provided weak empirical evidence to support the relationship between student achievement and teacher evaluation in elementary schools. Hallinger et al. (2014) concluded that the ideology of using VAMs was stronger than the actual evidence of its impact.

If value-added measures are to be used successfully in practice to recognize effective teachers, it is important to have a high level of confidence in the credit of achievement gains to specific teachers (Corcoran, 2010). If students are randomly assigned to teachers, it becomes

easier to determine that any differences between classroom achievement gains would mostly be due to the teacher. This would mean that other factors that influence year-to-year changes would essentially average out and uncover real differences in achievement gains across teachers. In reality, however, most students are not randomly assigned to classes, which is purposeful. Therefore, value-added methods must use a statistical model to determine the actual "value" a teacher adds to the classroom in relation to the students' academic growth.

Reliability Concerns Regarding Value-Added Models

Darling-Hammond et al.'s study (2012) examined teacher evaluation data from five school districts and found that 20% to 30% of the teachers who were rated less effective in any given year were rated the same the following year. Furthermore, 25% to 45% of the teachers who were rated less effective moved to the highly effective rating the following year. The same was true for those who were rated highly effective at the end of a year; namely, only a small minority stayed in the highly effective rating the following year. The researchers summarized three key limitations of using value-added measures for the purposes of teacher evaluation:

1. Value-added models of teacher effectiveness are inconsistent.
2. Teachers' value-added performance is affected by the students assigned to them, thereby calling into question the transparency and fairness of using value-added measures of student learning in evaluations.
3. Value-added ratings cannot disentangle the many influences on student progress and thereby provide an incomplete and inaccurate measure of an individual teacher's effectiveness. Most importantly, research has revealed that many more factors in addition to any individual teacher can influence gains in student achievement. These other factors include:

- School factors such as class sizes, curriculum materials, instructional time, availability of specialists and tutors, and resources for learning (books, computers, science labs, and more)
- Home and community supports or challenges
- Individual student needs and abilities, health, and attendance
- Peer culture and achievement
- Prior teachers and schooling, as well as other current teachers
- Differential summer learning loss, which especially affects low-income children
- The specific tests used, which emphasize some kinds of learning and not others, and which rarely measure achievement that is well above or below grade level (Darling-Hammond et al., 2012, pp. 2–4).

Many different VAMs have been used across the states although the largest and most widely implemented VAM in the country is the SAS Educational Value-Added Assessment System (SAS® EVAAS®). According to its developers, the SAS® EVAAS® is meant to "assess and predict student performance with precision and reliability" (Amrein-Beardsley & Collins, 2012, p. 15). The issue of reliability came into question from a study that Amrein-Beardsley and Collins completed (2012) in the Houston Independent School District (HISD). The authors examined the reliability of VAM outputs used as evidence to nonrenew four teachers. The study found several inconsistencies in the data among the four teachers. It was determined that VAM output data in three of the four teachers was unreliable, yet all four teachers were removed from their teaching positions due in part to their VAM scores.

Amrein-Beardsley and Collins (2012) pointed out that VAMs are sensitive and can fluctuate substantially within schools even when a different model is used or tested. Briggs and

Domingue conducted a similar study in 2011 that verified this observation. The study used an alternative statistical model to calculate the value-added scores for teachers in the Los Angeles Unified School District (LAUSD); these scores had been published in the *Los Angeles Times*. The results found that 40% to 55% of the teachers would receive different scores with the alternative model. For reading outcomes, 46% of teachers retained the same effectiveness rating under both models, 8.1% of teachers identified as effective under the alternative model were identified as more effective in LAUSD, and 12.6% of those identified as less or least effective under the alternative model were identified as relatively effective in the LAUSD model. For math outcomes, 60.8% of teachers retained the same effectiveness rating, 1.4 % of those teachers identified as effective under the alternative model were identified as ineffective in the LAUSD, and 2.7% would go from a rating of ineffective under the alternative model to effective under the LAUSD model.

Collins and Amrein-Beardsley (2014) compiled a study to capture state initiatives using growth models and to determine the strengths and weaknesses of each state's model. Approximately 40 states were in the process of using student growth models as part of their new teacher evaluation systems. The most popular VAMs used across the country were the EVAAS, the Student Growth Percentiles (SGP) model (also commonly known as the Colorado Growth Model [CGM]), the Value-Added Research Center (VARC) model, and homegrown models. In four states (including New Jersey), teacher consequences attached to growth or value-added data were locally controlled. In 15 states, teacher consequences attached to student performance data were yet to be determined, and in 14 states, teacher consequences would ultimately be attached to growth or value-added scores, which also influenced them. A total of 10 states tied or planned to tie teacher tenure decisions to value-added scores. As discussed in the research, controlling for

student characteristics such as socioeconomic status is important to the validity and reliability of VAMs. However, 21 states indicated that student characteristics were not accounted for in their growth model or VAM. Six states indicated that demographic information was accounted for, and nine states indicated that this information was yet to be determined.

In terms of reliability, some states expressed concerns with the accuracy of the data used when linking students to the teacher of record, and others expressed concerns on whether their state assessments were appropriately designed to measure teacher effectiveness over time. Darling-Hammond (2015) suggested:

> Standardized tests in the United States are criticized for their narrowness and focus on lower level skills; evidence has shown that high-stakes incentives to focus on these tests have reduced time spent teaching other important content and skills (Darling-Hammond & Adamson, 2014). Furthermore, because the NCLB Act mandated that state tests measure grade-level standards only, the tests do not include items that assess content or skills from earlier or later grade levels. As a result, these tests cannot measure the actual achievement level—or the learning gains—of the large share of students who are above or below grade level in their knowledge and skills (p. 132).

She found the same fault with more recently created national assessments (one of which was used in this study):

> The new tests created by the Partnership for Assessing Readiness for College and Careers (PARCC) and Smarter Balanced, the multistate consortia created to evaluate the Common Core State Standards, will not remedy this problem as they, too, have been required to measure grade-level standards. Even though they will

> report students' scores on a vertical scale, they will not be able to measure accurately the achievement or learning of students who started out below or above grade level (Darling-Hammond, 2015, p. 133).

Teacher Characteristics and the Impact on Student Achievement

The research findings related to teacher quality and its contribution to student achievement have been mixed. Some studies have found no or small effects of teacher characteristics, such as certification and experience, and several studies have attested that teachers contribute to student achievement.

Goldhaber and Brewer (1996) examined the test results of 18,000 students to estimate the impact of teacher degrees on student performance. The study found that several teacher characteristics were statistically significant and positively influenced student achievement. Teachers who were certified in mathematics and those with bachelor's or master's degrees in math and science were associated with higher student test scores. In another study, Goldhaber and Brewer (1999) examined teacher certification status and subject major and their relationships to student achievement using data from the National Educational Longitudinal Study of 1988. The authors found that students of teachers who had an undergraduate or graduate degree in mathematics performed better by a small margin of 0.08 standard deviation than students whose teachers did not have a mathematics degree. In addition, they found that students of teachers with any type of mathematics certification outperformed students whose teachers had no mathematics certification. These results suggested that subject knowledge of mathematics may be more important than the type of certification in terms of the contribution to student achievement.

In addition, Okpala, Smith, Jones, and Ellis (2000) conducted a study that determined a relative amount of significance for the impact of teacher characteristics on student achievement.

The population of the study consisted of 4,256 grade 4 students from 46 schools in North Carolina during the 1995–1996 school year. The two teacher characteristic variables that the study identified were the percent of teachers with master's degrees and percent of teachers with more than 10 years' teaching experience. The results from the study indicated that teachers with master's degrees explained achievement gains in mathematics with a significance level of 1 percent, but there was no significance in reading. Additional findings indicated a significant correlation between teachers with 10 years of teaching experience and student achievement in mathematics and reading with a significance level of 1 percent and 5 percent, respectively.

Goldhaber and Anthony (2007) examined the relationship between teachers with National Board for Professional Teaching Standards (NBPTS) certification and student achievement. It was determined that NBPTS is able to identify more effective teachers successfully among their applicants. In addition, National Board Certified Teachers (NBCT) were more effective than their noncertified peers at increasing student achievement before they became certified. The study matched 32,399 teachers to 609,160 students' reading test scores, and it matched 32,448 teachers to 611,517 mathematics test scores in North Carolina. The findings were marginally statistically significant:

> The magnitudes of the Future NBCT coefficients suggest that student gains produced by the teachers who are certified by NBPTS exceed those of noncertified applicants by about 4 percent of a standard deviation in reading and 5 percent of a standard deviation in math (based on a standard deviation of 9.94 on the end-of-year reading tests and 12.34 on the end-of-year math tests). These effects sizes are of the same order of magnitude as those found for math teachers having a bachelor's degree in their subject area (Goldhaber & Brewer, 1997). The

findings for Current NBCTs are smaller but still positive, and in the case of the reading model, statistically significant (p. 11).

Over time, several research studies have linked teacher classroom practices to student achievement. These practices include specific teaching strategies such as communicating clear learning objectives and expectations for student performance, utilizing standards-based learning objectives and assessments, and utilizing best instructional practices. Milanowski (2004) conducted a study in Cincinnati that analyzed the relationship between teacher evaluation scores and student achievement. The sample included 212 teacher evaluation scores using Danielson's (1996) framework and students in grades 3–8. The researcher found small to moderate correlations between teacher evaluation scores and student growth. The average correlations were 0.27 in science, 0.32 in reading, and 0.43 in mathematics.

School Characteristics and the Impact on Student Achievement

Kannapel and Clements (2005) examined 26 high-poverty elementary schools in Kentucky to determine what made high-performing, high-poverty schools different from other high-poverty schools. They selected eight high schools based on a school audit instrument that the state developed. All eight schools received high ratings in areas such as school culture and student, family, and community support. When these schools were compared with low-performing, high-poverty schools, significant findings were reported in a few areas. Teachers in the high-performing, high-poverty schools conducted frequent assessments, provided feedback to students, delivered instruction aligned to learning goals and assessments, had high expectations for student performance, used student achievement data for staff development purposes, and participated in collaborative decision making and job-embedded professional development.

Some research has shown that students' absenteeism rates can be detrimental to their success as early as kindergarten. In New Jersey, students who are absent for at least 10% of the school year are categorized as chronically absent. Gottfried (2014) examined the effect of chronic absenteeism on both achievement and socioemotional outcomes for a nationally representative dataset of kindergarten students from the 2010–2011 school year. The study found that chronic absenteeism could, in fact, reduce math and reading achievement outcomes and reduce educational engagement in schools.

Summary

Overall, the literature has provided evidence that classroom teachers have an impact on student achievement, and that the academic achievement of a student may depend on the quality or effectiveness of the teacher's instruction. Although some studies indicated value in the concept of holding teachers accountable for student learning, there was still no consensus on how to establish fair and equitable standards to evaluate teachers on their effectiveness and the value they add to a classroom. This lack of consensus is partly due to the incapability of VAMs to measure teacher qualities that could contribute to student learning and would be able to account for teacher evaluations. However, some findings have supported the use of VAMs to measure teacher effectiveness by controlling for factors that are outside the teacher's influence by using prior-year assessment data compared against the current year to measure the value of learning added during the year, which has created a national push to incorporate VAMs into teacher evaluations. The literature has identified the limitations in the use of VAMs to measure teacher effectiveness for the purposes of retention, promotion, compensation, and tenure.

The use of high-stakes tests to determine accountability measures using VAMs could manipulate teaching habits and influence teachers to teach to the test, even if it means changing

course from a preset curriculum guide. The literature has identified how VAMs may, in fact, not detect the effectiveness of teachers when teaching either low-performing or high-performing students. In many instances, the use of VAMs may discourage teachers from working in low-performing schools or with high-needs students, which creates longer vacancies and makes those positions harder to fill with certified teachers. Darling-Hammond (2015) reported that:

- Teachers of grades in which English language learners (ELLs) are transitioned into mainstreamed classrooms are the least likely to show "added value."
- Teachers of large numbers of special education students in mainstreamed classrooms are also found to have lower "value-added" scores on average.
- Teachers of gifted students show little added value because their students are already near the top of the test score range.
- Ratings change considerably when teachers change grade levels, often from "ineffective" to "effective" and vice versa (Darling-Hammond et al., 2012, p. 12).

In addition to the sensitivity of high-stakes tests' effects on teacher evaluation outcomes, the research literature has highlighted a range of issues related to the validity and reliability of VAMs. Some of these concerns stem from the fluctuation of evaluation ratings that may occur when teachers switch schools or change grade levels, which raises many questions regarding the validity of state tests and their ability to measure teacher effectiveness accurately. VAMs that do not control for student-level variables such as socioeconomics run the risk of masking the true effects of teachers on student achievement. VAMs do not specifically address the differences in instructional practices, use of questioning, and classroom management that have an effect on

student achievement between high- and low-performing teachers. They do not consider some of the most vital components that impact student learning such as chronic absenteeism, class sizes, curriculum materials, instructional time, or home and community supports. That being said, many states across the country have developed growth models and VAMs that help measure teacher effectiveness. For there to be a greater amount of trust that VAMs can inform teacher practice and appropriately measure teacher effectiveness, it is necessary for there to be more dialogue and research in this area.

The literature, however, has suggested a few areas in which certain teacher and school characteristics positively impact student achievement. Teachers who have specific qualifications in mathematics have been positively associated with increased student achievement. Students of teachers with stronger mathematics knowledge and mathematics certifications and degrees perform better than students of less knowledgeable mathematics teachers. In addition, teacher experience has demonstrated a positive impact on student achievement for the first few years of teaching. Measuring student achievement and a teacher's performance based on school-level characteristics also have an impact. The research has illustrated how teacher effectiveness can vary across schools and districts when some teachers work in challenging schools with many at-risk students, and others work in high-achieving schools in affluent suburban districts. A teacher rated effective in an affluent district may not be rated effective in a school with at-risk students in an urban setting. Therefore, both teacher and school characteristics play an important role in accurately measuring teacher effectiveness and can contribute to student success.

CHAPTER III

RESEARCH METHODOLOGY

Introduction

The method in which an administrator evaluates a teacher took a major shift in 2012 when the state of New Jersey enacted the TEACHNJ Act and AchieveNJ to reform teacher tenure laws and to link teacher tenure to evaluation ratings. Under AchieveNJ, multiple measures would now be used to evaluate teachers with the approach that students enter a school year at different levels of achievement, and that teachers should be credited for student improvement by integrating multiple measures of student growth into their evaluations. One of the essential questions in this study determined the value that teacher practices, teacher characteristics, and school characteristics added to student growth.

This study helps district leaders evaluate the new state mandates put in place for teacher evaluations and will contribute to the body of research related to teacher practice, specifically in an urban environment with large numbers of poor and low-performing students. The 2014–2015 school year was the 1st year that PARCC was administered in New Jersey and used to determine student growth. The 2017–2018 school year was the 4th year that PARCC was administered in New Jersey and used to determine student growth. It is also the 5th year that the state has implemented the new evaluation mandate, which allows this study to find the following: a) student achievement (growth) and its relationship to teacher characteristics; b) student achievement (growth) and its relationship to school characteristics; c) student achievement (growth) and its relationship to teacher practice, teacher characteristics, and school characteristics; and d) the comparison between student achievement (growth) and its relationship to teacher practice during the 2014-2015 and the 2017-2018 school year.

The 2014-2015 school year was the 2nd full year that the state of New Jersey provided SGP scores for students in the content areas of mathematics and language arts; the 2017–2018 school year was the 5th full year of SGP scores in New Jersey. An SGP describes students' growth relative to their academic peers who had the same PARCC scores for the past 3 years. Students are measured against their peers to determine academic growth and are provided an SGP score categorized as low, typical, or high growth. For the purpose of this study, the students' SGP was the dependent variable used to measure student growth.

This chapter discusses the methods and procedures used to examine the relationship between teacher practice and student performance in a large urban New Jersey school district. The methods and procedures are discussed in the following sections: a) Methods, b) Design, c) Participants, d) Setting, e) Instrumentation and Variables, f) Procedures, and g) Data Analysis.

Methods

This study used a quantitative methodology, which is the most appropriate format to answer the research questions. Quantitative research is used to answer questions about the relationships between measured variables and quantify a problem by generating numerical data that can be transformed into usable statistics. In many cases, quantitative research will use longitudinal data to examine trends and will often include structural equation models that identify strengths in multiple variables (Creswell & Creswell, 2017). This study sought to explain the relationships between variables that predict student academic growth. The study used SGP scores, which schools receive from the state annually, as the dependent variable to run its analysis. The study also used teacher- and school-level data available from the district studied.

Design

This study used a cross-sectional explanatory design to explain how teacher practice scores predict student growth. The data for this study were captured in 2 separate years. First, the study identified relationships during the 2014-2015 school year. Next, the study identified relationships during the 2017-2018 school year and made comparisons between the 2 years. This study began with the 2nd full year of implementation for the AchieveNJ mandate in which student SGP scores (student growth) in language arts and mathematics were calculated and reported as a multiple measure for teacher evaluation purposes. Furthermore, this study began with the 1st year that the New Jersey Department of Education administered the standardized statewide PARCC assessments in language arts and mathematics. This study examined the relationships between teacher characteristics, school-level variables, and teacher practice scores on student academic growth.

Participants

The population identified in this study consisted of students with a valid SGP score in grades 4–7 in the content areas of language arts or mathematics in both school years. The study consisted of 12 schools in 2014-2015 and 14 schools in 2017-2018 with different grade configurations as indicated in Table 2.

Table 2

Grade Configurations of Schools with Grades 4, 5, 6, and/or 7

Grade Configuration	K-5	PK-6	PK-7	PK-8	6-8	6-12	Total
Number of schools (2014-2015)	0	1	1	7	2	1	12
Number of schools (2017-2018)	1	1	1	9	0	2	14

The New Jersey Department of Education categorized schools within districts as "Needs Comprehensive Support" based on schoolwide proficiency and growth levels that the state assessment, PARCC, measured along with chronic absenteeism rates. Schools are identified for comprehensive support and improvement if they meet one of the three following criteria: a) A school's overall performance is at or below the 5th percentile of Title I schools, b) a high school has a 4-year graduation rate at or below 67 percent, and c) a Title I school is identified as in need of targeted support and improvement for 3 or more consecutive years (i.e., the school has a chronically low-performing subgroup; NJ Department of Education, 2018). Other schools are identified for targeted support and improvement if a school has a student subgroup whose overall performance is at or below the 5th percentile of Title I schools (NJ Department of Education, 2017). For this study, schools that were in need of comprehensive support were listed as "Comprehensive", and schools that were in need of targeted support and improvement were listed as "Target". Schools that were not in need of either comprehensive or targeted support were listed as "NonStatus". For the purposes of this study, schools in the 2014-2015 school year that were labeled Priority were listed as Comprehensive, and schools that were labeled as Focus were listed as Target. In this study, all of the schools in the 2014-2015 school year were labeled as Comprehensive. However, in the 2017-2018 school year there were six Comprehensive schools, four Target schools, and four NonStatus schools.

The sample population for the 2014-2015 school year consisted of 1,132 students (n = 1,132) with a language arts SGP and 1,087 students (n = 1,087) with a mathematics SGP in grades 4–7. The sample population for the 2017-2018 school year consisted of 1,484 students (n = 1,484) with a language arts SGP and 1,473 students (n = 1,473) with a mathematics SGP in grades 4–7. Each student in the study had received a valid SGP score during that school year.

The study will exclude bilingual teachers because there were only a few valid SGP scores for those teachers.

Setting

The study will take place in a large urban school district that enrolls roughly 6,800 students from preschool through grade 12. The district's population is 48% Hispanic, 48% African American, and 4% Caucasian, Middle Eastern, or Asian descent. The district consists of 18 schools with approximately 766 certified teachers. The New Jersey Department of Education (NJDOE) identified several schools in the district as schools "in need of comprehensive or targeted support and improvement." It is one of the 31 former Abbott school districts and one of two school districts that are under state control.

The study followed the mandates that the new evaluation policy (AchieveNJ) set forth, which were fully implemented for the first time during the 2013–2014 school year. In compliance with the mandate, the district evaluated approximately 766 teachers using a state-approved evaluation instrument named Charlotte Danielson: The Framework for Teaching (2013 Edition). Substantial professional development on the new evaluation system is provided for teachers and administrators on an annual basis. During the school year, school leaders use their biweekly professional development time to work on strategies to strengthen teacher practice.

Other than teacher practice, the new teacher evaluation policy measures student achievement using SGP scores based on PARCC test results. For this study, SGP data was only available for students in grades 4–7 in language arts and mathematics. Grade 8 students were omitted from the study because the state does not calculate SGP in grade 8 mathematics.

Instrumentation and Variables

The PARCC state assessment was administered for the first time in New Jersey in the spring of 2015 and meets the reliability and validity criteria as indicated in the Partnership for Assessment of Readiness for College and Careers, 2017, Final Technical Report (PARCC Technical Report, 2018):

> The test reliabilities measured by average reliability estimates for the 2017 PARCC are described in Section 8. The average reliability estimates for overall student scaled scores in grades 3-8 ranged from .962 to .966 on Computer-Based Tests (CBTs) and .958 to .966 on Paper-Based Tests (PBTs) for ELA, and from .919 to .943 on CBTs and .909 to .944 on PBTs for mathematics, indicating that the tests are highly reliable (p. 99).

Test validity is reflected in a process where:

> Analyses of the internal structure of a test typically involve studies of the relationships among test items and/or test components (i.e., subclaims) in the interest of establishing the degree to which the items or components appear to reflect the construct on which a test score interpretation is based. The term construct is used here to refer to the characteristics that a test is intended to measure; in the case of the PARCC operational tests, the characteristics of interest are the knowledge and skills defined by the test blueprint for ELA/L and for mathematics.
>
> The PARCC assessments provide a full summative test score, Reading claim score, and Writing claim score as well as ELA/L subclaims and mathematics subclaim scores. The goal of reporting at this level is to provide criterion-

> referenced data to assess the strengths and weaknesses of a student's achievement in specific components of each content area. This information can then be used by teachers to plan for further instruction, to plan for curriculum development, and to report progress to parents. The results can also be used as one factor in making administrative decisions about program effectiveness, teacher effectiveness, class grouping, and needs assessment (PARCC Technical Report, 2018, p. 137).

The manual explained:

> The PARCC item development process involved educators, assessment experts, and bias and sensitivity experts in review of text, items, and tasks for accuracy, appropriateness, and freedom from bias. PARCC conducted several studies during the item development process to evaluate the item development process (e.g., technological functionalities, answer time required, and student experiences). The intercorrelations of the subclaims, the reliability analyses, and the local item dependence analyses indicated that the ELA/L and the mathematics assessments are both essentially unidimensional. Also, the patterns of correlations for the CBT and PBT assessments were similar, indicating that the structures of the assessments were similar across the two modes. The benchmarking study was conducted in support of the standard setting meeting. This study indicated that students performing at or above Level 4 could be considered to be college- and career-ready or on track to readiness. The content evaluation/alignment studies performed by the Fordham Institute and HumRRO indicate that the PARCC assessments are good to excellent matches to the CCSS in terms of content and

> depth of knowledge. Thus, the PARCC assessments are assessing the college- and career-readiness standards. In the longitudinal study of external validity, associations between PARCC performance levels and college-readiness benchmarks established by the College Board and ACT were used to study the claim that students who achieve Level 4 have a .75 probability of attaining at least a C in entry-level, credit bearing, postsecondary coursework. In the first phase of the study, the relationship between PARCC and external tests was studied. Overall, results indicated that a student meeting the benchmark on the PARCC test had a high probability of making the benchmark on the external test, but the converse did not hold for students meeting the benchmark on the external test, for the majority of comparisons. These results suggest that meeting the PARCC benchmark is an indicator of academic readiness for college (PARCC Technical Report, 2018, p. 169).

The mathematics portion of PARCC measured a student's ability to apply skills and concepts, through both short- and extended-response questions, to solve problems that demonstrate knowledge of mathematical practices from the CCSS with a focus on modeling and reasoning with precision. The mathematics test also consisted of performance-based short-answer questions focused on conceptual understanding, procedural skills, and application (PARCC Technical Report, 2018).

The ELA/L PARCC assessment measured students' literary analysis, research simulation, and narrative writing through either literacy and informational reading passages or performance-based tasks. The reading passages would include a combination of both vocabulary and comprehension questions. Each of the performance-based tasks would ask students to read or

view one or more texts, answer comprehension and vocabulary questions, and write an extended response that would require them to draw evidence from the text(s; PARCC Technical Report, 2018).

A variety of data collection instruments were used in this study to determine empirically the relationship between teacher practice and student achievement. The following instruments were used: a) student SGP scores, b) teacher-level characteristics, c) school-level characteristics, and d) teacher evaluation practice scores.

Student Growth Percentile (SGP) Scores

Student growth was measured by using the student's SGP score that the state's Department of Education provided for each teacher in grades 4–7 in language arts and mathematics. An SGP describes students' growth relative to their academic peers who had the same state assessment (PARCC) scores for, at most, the past 3 years when applicable. Betebenner (2011) explained:

> If the student's current year score exceeded the scores of most of their academic peers, in a normative sense they have done well. If the student's current year score was less than the scores of their academic peers, in a normative sense they have not done well (p. 3).

The change in student growth is reported as an SGP and specified on a scale from 1 to 99 to rate how students' growth compared to their academic peers. It then categorizes each student as having either low growth (SGP < 35), typical growth (SGP > 34 and SGP < 66), or high growth (SGP > 65).

Teacher Characteristics

Model 1 used a logistic regression analysis to test the relationship between the predictor variable teacher characteristics and student growth (dependent variable) to determine if teacher characteristics such as gender, ethnic background, and age were significant predictors of student growth. Each teacher characteristic was coded as follows:

gender	0 = female, 1 = male
ethnicity	0 = Black, 1 = all others
	0 = Hispanic, 1 = all others
	0 = White/Asian, 1 = all others
age	scale variable

School Characteristics

Model 2 used a logistic regression to test the relationship between the predictor variable, school characteristics, and student growth (dependent variable) to determine to what degree school-level variables such as school performance status (Comprehensive, Target, or NonStatus schools) and the percent of student subgroup ethnic composition influenced student growth. School characteristics were coded based on school performance status:

0 = Comprehensive, 1 = Target and NonStatus

0 = Target, 1 = Comprehensive and NonStatus

0 = NonStatus, 1 = Comprehensive and Target

The ethnic subgroup characteristics addressed the percentage of each ethnic subgroup enrolled in each school:

Black enrollment	0 = 69% and less than, 1 = 70% and greater than
Hispanic enrollment	0 = 69% and less than, 1 = 70% and greater than
White/Asian enrollment	0 = 2% and less than, 1 = 3% and greater than

Teacher Evaluation Practice Score

The teacher evaluation practice score was used as an independent variable in a logistic regression model to determine the value added to student growth. A teacher practice score was derived from an average of observation ratings by components and domains, and weighted as specified by the Charlotte Danielson: The Framework for Teaching Model (2013 Edition), which the district adopted and the state's Department of Education approved to observe teacher practice. The framework focuses on the following four performance criteria for teachers: a) planning and preparation, b) classroom environment, c) instruction, and d) professional responsibilities. Teacher practice was represented as a score from 1 (ineffective) to 4 (highly effective). For this study, teacher practice was categorized as a scale variable from 1 (ineffective) to 4 (highly effective). Table 3 lists the variables and measurements that were conducted in the study.

Table 3

Instrumentation and Variables

Variables	Measurement	Status
SGP (Student Growth)	SGP: 0 = Low Growth 1 = Typical and High Growth	Dependent
Teacher Characteristics	Gender: 0 = Male 1 = Female Ethnicity: 0 = Black 1 = All others 0 = Hispanic 1 = All others 0 = White/Asian 1 = All others Age: Scale (min = 24 years old, max = 68 years old)	Independent
School Characteristics	School performance status: 0 = Comprehensive 1 = Target and NonStatus School performance status: 0 = Target 1 = Comprehensive and NonStatus School performance status: 0 = NonStatus 1 = Comprehensive and Target Black enrollment percentage: 0 = 69% and less than 1 = 70% and greater than Hispanic enrollment percentage: 0 = 69% and less than 1 = 70% and greater than White/Asian enrollment percentage: 0 = 2% and less than 1 = 3% and greater than	Independent
Teacher Practice	Scale (min = 1.00, max = 4.00)	Independent

Procedures

The district has granted authorization to conduct the research study. The process for approval included a *District Data/Research Request Form* that was completed by answering 13 questions pertaining to the study. The Department of Education's NJ Standards Measurement and Resource for Teaching (NJSMART) is a comprehensive statewide longitudinal data system that serves multiple purposes and includes teacher/student identification, data warehousing, data reporting, and analytics. The data on which the findings were based were collected from an NJSMART file.

Evidence of teacher practice scores was gathered from the district's reporting system that was uploaded into the state's NJSMART portal. The state calculated and provided the SGP scores for individual students. A comprehensive data file was downloaded from the NJSMART portal that included the teachers' practice scores, students' SGP scores, student-assigned schools, and teacher-assigned students. The district provided a data file that identified each teacher's characteristics in terms of gender, ethnicity, and age. School characteristics were collected from the New Jersey Department of Education website that categorizes schools in this study as Comprehensive or Target schools. Schools not considered as Comprehensive or Target schools were regarded as NonStatus schools in this study. The school performance reports on the New Jersey Department of Education website provided each school's student ethnicity enrollment.

Data Analysis

In this study, fixed effects logistic and hierarchical fixed effects logistic regression methods were used to test the relationships between the dependent variable (student growth) and independent variables (teacher characteristics, school characteristics, teacher practice) in two different time periods. This logistic regression method allowed the researcher to identify the

magnitude of independent variables (gender, ethnicity, age, school performance, and student ethnicity) that contributed to student growth. The hierarchical logistic method was used to better explain the relationship between the independent (teacher practice, teacher characteristics, school characteristics) and dependent (student growth) variables. Specifically, the model explained the probability of students' demonstrating typical or high growth if they had a teacher with a high practice score.

The design consisted of three separate models used to answer four research questions. A fixed effects logistic regression analysis was used to analyze Model 1 (teacher characteristics on student growth) and Model 2 (school characteristics on student growth). In Model 3, hierarchical fixed effects logistic regression analysis was used to better interpret the impact of teacher practice and teacher and school characteristics on student growth. The dependent variable SGP dummy coded variables (low, typical, and high) were collapsed into two variables (low and typical/high), and the predictor variables from Models 1 and 2 were included in the model with teacher practice. Model 4 compared the findings in Model 3 from the 2014-2015 to the 2017-2018 school year.

Model 1 determined the extent to which the teacher characteristics of gender, ethnicity, and age could predict the odds of typical or high growth in the outcome variables of LAL SGP and Math SGP. Gender and ethnicity were dichotomized, and age was used as a scale variable. Furthermore, ethnicity was separated into three teacher ethnicity groups: Black, Hispanic, and White or Asian. Model 2 determined the extent to which the school characteristics of comprehensive status, target status, and student ethnicity enrollment subgroups could predict the odds of typical or high growth in the outcome variables of LAL SGP and Math SGP. Each of the school independent variables was dichotomized. Furthermore, student ethnicity enrollment was

separated into three ethnic subgroups: Black student enrollment percentage, Hispanic student enrollment percentage, and White or Asian student enrollment percentage. Model 3 added teacher practice scores, main predictor variable, to the hierarchical logistic regression, which included the teacher and school characteristics. Teacher Practice Score was used as a scale variable and determined how much value it added to predict the outcome variables of LAL SGP and Math SGP.

This analysis identified the value that teacher practice and teacher and school characteristics added to student growth during 2 separate years to distinguish any improvements in their relationship. The model presented findings on student growth from multiple angles, specifically on how teacher practice, teacher gender, teacher ethnicity, teacher age, school-level performance, and school ethnic subgroups relate to student growth. Variables were coded to distinguish differences in characteristics.

Preliminary analysis was conducted to examine the "goodness-of-fit" in each model according to the Hosmer and Lemeshow Test. "The purpose of any overall goodness-of-fit test is to determine whether the fitted model adequately describes the observed outcome experience in the data" (Archer & Lemeshow, 2006, p. 97). Both the Cox and Snell and Nagelkerke R Squares were used to determine the percentage of the variance in the outcome variable SGP that the predictor variables could explain. In Model 3, the Nagelkerke R Square was used to describe the increase in the explanation of the variance in the outcome variable SGP after adding the main predictor variable, teacher practice, to the overall model. A $p < .05$ level of significance was used for all analyses in the study to determine if a variable had significance in explaining the outcome variable SGP.

Analyses were also conducted to test the predictive value of the logistic model (Model 3). The classification tables for the samples of students' growth level in language arts and mathematics during the 2014-2015 and 2017-2018 school years were examined and provided a measure of the discriminative efficiency of the logistic regression model. SPSS was used to determine the degree of variance among the models with logistic regression analyses. A detailed analysis of the data, the findings, and conclusions are presented in Chapters IV and V.

CHAPTER IV

RESEARCH FINDINGS

Introduction

The central purpose of this study was to understand the relationship between teacher practice and student growth in a large urban district. For many years, educational accountability to improve student achievement on standardized tests was the topic of conversation among education reformers. Due to the Race to the Top grant program, states were incentivized for creating innovative education reform that could demonstrate the ability to make significant growth in student outcomes by closing the achievement gap and redesigning their evaluation systems (U.S. Department of Education, 2009). New Jersey was one of the states awarded this grant and began the work to improve educator evaluations to increase teacher and leader effectiveness.

These new evaluation policies were a major shift from prior evaluation procedures. The policies eliminated the binary method of evaluation (i.e., satisfactory vs. unsatisfactory), which did not take student growth into account. The new system, however, now incorporates student growth along with a teacher's practice to produce an overall performance score that reflects one of four different teacher rating categories (highly effective, effective, partially effective, and ineffective). Sanders (2000) pointed out that a fair accountability system should measure teacher effectiveness by the rate of student progress regardless of socioeconomic status, and that such a system would need to acknowledge that all students are at different academic levels and will learn at different paces.

The present research represented a cross-sectional explanatory design in which fixed effects logistic and hierarchical fixed effects logistic regressions were used to answer questions

related to how teacher practice predicts student growth. The analyses consisted of testing three models for the first three research questions, which examined the relationship between student academic growth while controlling for teacher practice and teacher- and school-level characteristics. The fourth question was a comparison analysis between the 2014-2015 and 2017-2018 school years in the third model. This chapter presents the findings from these analyses.

The research was guided by the following questions:

1. Are teacher-level variables such as gender, ethnic background, and age significant predictors of student growth?
2. To what extent do the following school-level variables influence student growth: school performance status (Comprehensive schools, Target schools, and NonStatus schools) and percent of student subgroup ethnic composition?
3. How is student growth in language arts and mathematics impacted by a teacher's effectiveness as measured by the practice score received when one controls for teacher- and school-level characteristics?
4. To what extent does the relationship between teacher effectiveness and student outcome in language arts and mathematics vary from year 2 of AchieveNJ to year 5 of AchieveNJ?

This chapter presents the major findings. The outcome variable (student growth) was operationalized using Student Growth Percentile (SGP) categorized as low, typical, or high growth (NJ SMART Education Data System, 2016).

Participant Demographics

Table 4 and Table 5 provide a breakdown of the teacher characteristics of the 1,161 students in the 2014-2015 school year with valid SGP scores in either language arts or

mathematics included in the study. Of 736 students with a valid language arts SGP, 15% had a male language arts teacher, and 85% had a female language arts teacher; further, of 784 students with a valid mathematics SGP, 19% had a male mathematics teacher, and 81% had a female mathematics teacher. In addition, of 736 students with a valid language arts SGP, 66% of their teachers were Black, and 34% were White; of 784 students with a valid mathematics SGP, 51% of their teachers were Black, and 49% were White. Lastly, the average age of the language arts teachers in the study was 46 years old with a standard deviation of 9 years; the average age of the mathematics teachers in the study was 44 years old with a standard deviation of 9 years.

Table 4

2014-2015 Teacher Characteristics: Gender, Ethnicity

Teacher Characteristics		Frequency	Percentage
LAL Teacher Gender 0 = Male; 1 = Female	Students with Male Teacher	110	14.9%
	Students with Female Teacher	626	85.1%
Math Teacher Gender 0 = Male; 1 = Female	Students with Male Teacher	152	19.4%
	Students with Female Teacher	632	80.6%
LAL Teacher Black Ethnicity 0 = Black; 1 = All Others	Students with Black Teacher	484	65.8%
	Students with Other Teacher	252	34.2%
LAL Teacher White Ethnicity 0 = White; 1 = All Others	Students with White Teacher	252	34.2%
	Students with Other Teacher	484	65.8%
Math Teacher Black Ethnicity 0 = Black; 1 = All Others	Students with Black Teacher	397	50.6%
	Students with Other Teacher	387	49.4%
Math Teacher White Ethnicity 0 = White; 1 = All Others	Students with White Teacher	387	49.4%
	Students with Other Teacher	397	50.6%

Table 5

2014-2015 Teacher Characteristics: Age

Teacher Characteristics	N	Min	Max	Mean	Std. Deviation
LAL Teacher Age	736	24 years	65 years	46.38 years	8.54 years
Math Teacher Age	784	24 years	65 years	44.46 years	8.97 years

Table 6 and Table 7 provide a breakdown of the teacher characteristics of the 1502 students in the 2017-2018 school year with valid SGP scores in either language arts or mathematics included in the study. Of 1,484 students with a valid language arts SGP, 13% had a male language arts teacher, and 87% had a female language arts teacher; further, of 1,438 students with a valid mathematics SGP, 16% had a male mathematics teacher, and 84% had a female mathematics teacher. In addition, of 1,484 students with a valid language arts SGP, 59% of their teachers were Black, 37% were White or Asian, and 4% were Hispanic; of 1,438 students with a valid mathematics SGP, 50% of their teachers were White or Asian, 47% were Black, and 3% were Hispanic. Lastly, the average age of the language arts teachers in the study was 48 years old with a standard deviation of 10 years; the average age of the mathematics teachers in the study was 48 years old with a standard deviation of 8 years.

Table 6

2017-2018 Teacher Characteristics: Gender, Ethnicity

Teacher Characteristics		Frequency	Percentage
LAL Teacher Gender 0 = Male; 1 = Female	Students with Male Teacher	189	12.7%
	Students with Female Teacher	1,295	87.3%
Math Teacher Gender 0 = Male; 1 = Female	Students with Male Teacher	229	15.9%
	Students with Female Teacher	1,209	84.1%
LAL Teacher Black Ethnicity 0 = Black; 1 = All Others	Students with Black Teacher	869	58.6%
	Students with Other Teacher	615	41.4%
LAL Teacher White/Asian Ethnicity 0 = White/Asian; 1 = All Others	Students with White/Asian Teacher	549	37.0%
	Students with Other Teacher	935	63.0%
LAL Teacher Hispanic Ethnicity 0 = Hispanic; 1 = All Others	Students with Hispanic Teacher	66	4.4%
	Students with Other Teacher	1,418	95.6%
Math Teacher Black Ethnicity 0 = Black; 1 = All Others	Students with Black Teacher	678	47.1%
	Students with Other Teacher	760	52.9%
Math Teacher White/Asian Ethnicity 0 = White/Asian; 1 = All Others	Students with White/Asian Teacher	721	50.1%
	Students with Other Teacher	717	49.9%
Math Teacher Hispanic Ethnicity 0 = Hispanic; 1 = All Others	Students with Hispanic Teacher	39	2.7%
	Students with Other Teacher	1,399	97.3%

Table 7

2017-2018 Teacher Characteristics: Age

Teacher Characteristics	N	Min	Max	Mean	Std. Deviation
LAL Teacher Age	1,484	25 years	68 years	47.97 years	10.17 years
Math Teacher Age	1,438	25 years	68 years	48.34 years	8.32 years

The school-level variables for the 2014-2015 school year included in the study are presented in Table 8. There were 12 schools in the study, and every school was listed as a Comprehensive school. The students were enrolled in schools with varying student populations. Of the 1,161 students with a valid SGP, 874 students were enrolled in schools with a student subgroup African-American population of less than 70%, 1,039 students were enrolled in schools with a Hispanic subpopulation of less than 70%, and 822 students were enrolled in schools with a White or Asian student population of less than 3%.

Table 8

2014-2015 School Characteristics: School Performance Designation, Student Ethnicity

School Characteristics		Frequency	Percentage
School Comprehensive Status 0 = Comprehensive; 1 = Target and NonStatus	Comprehensive	1,161	100%
	Target and NonStatus	0	0.0%
School Black Enrollment Racial Subgroup 0 = less than 70%; 1 = greater than 69%	69% and less	874	75.3%
	70% and greater	287	24.7%
School Hispanic Enrollment Racial Subgroup 0 = less than 70%; 1 = greater than 69%	69% and less	1,039	89.5%
	70% and greater	122	10.5%
School Asian/White Enrollment Racial Subgroup 0 = less than 3%; 1 = greater than 2%	2% and less	822	70.8%
	3% and greater	339	29.2%

The school-level variables for the 2017-2018 school year included in the study are presented in Table 9. There were 14 schools: 6 Comprehensive schools, 4 Target schools, and 4 NonStatus schools. In terms of student placement, 51% of the students were from Comprehensive schools, 20% were from Target schools, and 29% were from NonStatus schools. The students were enrolled in schools with varying student populations. Of the 1,502 students

with a valid SGP, 1,180 students were enrolled in schools with a student subgroup African-American population of less than 70%, 975 students were enrolled in schools with a Hispanic subpopulation of less than 70%, and 1,043 students were enrolled in schools with a White or Asian student population of less than 3%.

Table 9

2017-2018 School Characteristics: School Performance Designation, Student Ethnicity

School Characteristics		Frequency	Percentage
School Comprehensive Status 0 = Comprehensive; 1 = Target and NonStatus	Comprehensive	769	51.2%
	Target and NonStatus	733	48.8%
School Target Status 0 = Target; 1 = Comprehensive and NonStatus	Target	296	19.7%
	Comprehensive and NonStatus	1,206	80.3%
School NonStatus Status 0 = NonStatus; 1 = Comprehensive and Target	NonStatus	437	29.1%
	Comprehensive and Target	1,065	70.9%
School Black Enrollment Racial Subgroup 0 = less than 70%; 1 = greater than 69%	69% and less	1,180	78.6%
	70% and greater	322	21.4%
School Hispanic Enrollment Racial Subgroup 0 = less than 70%; 1 = greater than 69%	69% and less	975	64.9%
	70% and greater	527	35.1%
School Asian/White Enrollment Racial Subgroup 0 = less than 3%; 1 = greater than 2%	2% and less	1,043	69.4%
	3% and greater	459	30.6%

Table 10 and Table 11 provide a breakdown of both student and teacher performance for the 2014-2015 school year. Regarding student performance, the findings indicate that 53% of students had low growth in language arts, and 56% had low growth in mathematics; 47% of

students had typical or high growth in language arts, and 44% had typical or high growth in mathematics.

Regarding teacher performance for the 2014-2015 school year, the findings indicate that teachers with an effective practice score in either language arts or mathematics instructed most students. Specifically, although the minimum practice score of a language arts teacher was 1.70 (ineffective), the average practice score was 2.92 (effective) with a standard deviation of .40, which means that approximately 68% of language arts teachers had a practice score between 2.52 (partially effective) and 3.32 (effective). In addition, although the minimum practice score of a mathematics teacher was 1.70 (ineffective), the average practice score was 2.96 (effective) with a standard deviation of .35, which means that approximately 68% of language arts teachers had an effective practice score between 2.61 (partially effective) and 3.31 (effective).

Table 10

2014-2015 Teacher Evaluation Rating: Student Growth Percentile (SGP)

Dependent Variable		Frequency	Percentage
LAL SGP Growth Level 0 = Low Growth; 1 = Typical or High Growth	Low Growth	600	53.0%
	Typical or High Growth	532	47.0%
Math SGP Growth Level 0 = Low Growth; 1 = Typical or High Growth	Low Growth	611	56.2%
	Typical or High Growth	476	43.8%

Table 11

2014-2015 Teacher Evaluation Rating: Teacher Practice

Teacher Practice	N	Min	Max	Mean	Std. Deviation
LAL Teacher Practice Score	1132	1.70	3.67	2.92	0.40
Math Teacher Practice Score	1087	1.70	3.67	2.96	0.35

Table 12 and Table 13 provide a breakdown of both student and teacher performance for the 2017-2018 school year. Regarding student performance, the findings indicate that 49% of students had low growth in language arts, and 47% had low growth in mathematics; 51% of students had typical or high growth in language arts, and 53% had typical or high growth in mathematics.

Regarding teacher practice scores, a teacher with an effective practice score in either language arts or mathematics instructed most students. Specifically, although the minimum practice score of a language arts teacher was 2.11 (partially effective), the average practice score was 3.13 (effective) with a standard deviation of .28, which means that approximately 68% of the language arts teachers had an effective practice score between 2.85 and 3.41. In addition, although the minimum practice score of a mathematics teacher was 2.35 (partially effective), the average practice score was 3.16 (effective) with a standard deviation of .27, which means that approximately 68% of the language arts teachers had an effective practice score between 2.89 and 3.43.

Table 12

2017-2018 Teacher Evaluation Rating: Student Growth Percentile (SGP)

Dependent Variable		Frequency	Percentage
LAL SGP Growth Level 0 = Low Growth; 1 = Typical or High Growth	Low Growth	725	48.9%
	Typical or High Growth	759	51.1%
Math SGP Growth Level 0 = Low Growth; 1 = Typical or High Growth	Low Growth	699	47.5%
	Typical or High Growth	774	52.5%

Table 13

2017-2018 Teacher Evaluation Rating: Teacher Practice

Teacher Practice	N	Min	Max	Mean	Std. Deviation
LAL Teacher Practice Score	1484	2.11	4.00	3.13	0.28
Math Teacher Practice Score	1473	2.35	4.00	3.16	0.27

Findings Model 1

Research Question 1 asked: Are teacher-level variables such as gender, ethnic background, and age significant predictors of student growth in both language arts and mathematics? Model 1 tested the relationship between teacher characteristics and student growth to determine if teacher-level variables such as gender, ethnic background, and age significantly explained students' growth in both language arts and mathematics.

Logistic regression models were run for both the 2014-2015 and 2017-2018 school years to determine the extent to which the predictor variables of Teacher Gender, Teacher Ethnicity, and Teacher Age can predict the odds of typical or high growth in the outcome variables of LAL SGP and Math SGP. To facilitate understanding of the findings, the dummy codes for gender and ethnic background are listed below.

Teacher Characteristics:

gender: 0 = male 1 = female;

ethnic background: 0 = Black 1 = all others;

0 = Hispanic 1 = all others;

0 = White/Asian 1 = all others;

age: Scale

The New Jersey Department of Education categorizes the outcome variable (student growth) as low, typical, or high growth. The outcome variable in the study was dichotomized as low growth and typical/high growth.

Findings 2014-2015

During the 2014-2015 school year, the findings in Table 14 indicate that the teacher characteristics in the study have a low "fitness" level in both language arts and mathematics according to the Hosmer and Lemeshow Test (Chi Square = 21.512, df = 7, $p < .003$ for language arts and Chi Square = 58.741, df = 8, $p < .000$ for mathematics). In addition, Table 15 presents the model summary statistics, which indicated that the teacher characteristic predictors can explain only between 0.9% and 1.2% of the variance in the outcome variable of LAL SGP; the teacher characteristic predictors can explain only between 0.4% and 0.5% of the variance in the outcome variable of Math SGP.

In the logistic regression model for predicting typical or high growth in language arts, only teacher gender had a statistically significant relationship to growth. Table 16 indicates that with respect to the gender of a language arts teacher, the odds ratio of 1.671 was significant (Wald = 5.634, p < .018, [CI = 1.094 – 2.553]). The odds ratio of 1.671 indicates that students with female teachers were more likely to have typical or high growth than students with male teachers. Specifically, the probability that a student would have typical or high growth in language arts increased by 67% with a female teacher compared to students with a male teacher. In mathematics, none of the teacher demographic predictors were found to be significant.

Table 14

2014-2015 Model 1 Hosmer and Lemeshow Test

Model	Chi-square	df	Sig.*
Language Arts	21.512	7	.003
Mathematics	58.741	8	.000

Note: df = degrees of freedom
*$p > .05$

Table 15

2014-2015 Model 1 Model Summary

Model	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Language Arts	1013.421	.009	.012
Mathematics	1082.527	.004	.005

Table 16

2014-2015 Variables in the Equation: *Impact of Teacher Characteristics on Student Growth*

								95% Confidence Interval	
Model	N	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
LAL Teacher Gender	736	.513	.216	5.634	1	.018	1.671	1.094	2.553
LAL Teacher Black Ethnicity	736	.081	.158	.261	1	.610	1.084	.796	1.476
LAL Teacher Age	736	.001	.009	.014	1	.906	1.001	.984	1.018
Math Teacher Gender	784	.283	.189	2.231	1	.135	1.327	.916	1.922
Math Teacher Black Ethnicity	784	-.171	.149	1.312	1	.252	.843	.629	1.129
Math Teacher Age	784	-.002	.008	.081	1	.776	.998	.982	1.014

*$p < .05$

Findings 2017-2018

During the 2017-2018 school year, the findings in Table 17 indicate that the teacher characteristics in the study have a low "fitness" in both language arts and mathematics according to the Hosmer and Lemeshow Test (Chi Square = 15.829, df = 8, $p < .045$ for language arts and Chi Square = 28.62, df = 8, $p < .000$ for mathematics). In addition, Table 18 presents the model summary statistics, which indicated that the teacher characteristic predictors can explain only between 0.6% and 0.8% of the variance in the outcome variable of LAL SGP; the teacher characteristic predictors can explain only between 1.1% and 1.5% of the variance in the outcome variable of Math SGP.

In the logistic regression model for predicting typical or high growth in language arts, only teacher ethnicity had a statistically significant relationship to growth. Table 19 indicates that with respect to the ethnicity of a language arts teacher, the odds ratio of .776 for LAL Teacher Black Ethnicity was significant (Wald = 5.166, p < .023, [CI = .624 – .966]). The odds ratio of .776 indicates that students with a Black teacher were less likely to have typical or high growth than students with a teacher who was not Black. Specifically, the probability that a student would have typical or high growth in language arts decreased by 22% with a Black teacher compared to students with a non-Black teacher. With respect to the ethnicity of a mathematics teacher, the odds ratio of 3.524 for Math Teacher Hispanic Ethnicity was significant (Wald = 11.511, p < .001, [CI = 1.702 – 7.295]). The odds ratio of 3.524 indicates that students with a Hispanic teacher were more likely to have typical or high growth than students with a teacher who was not Hispanic. Specifically, the probability that a student would have typical or high growth in mathematics increased by 352% with a Hispanic teacher compared to students with a non-Hispanic teacher. In addition, in the logistic regression model for predicting typical or high

growth in mathematics, teacher age had a statistically significant relationship to growth. The odds ratio of .988 for Math Teacher Age was significant (Wald .081, $p < .052$, [CI = .975 – 1.000]). The odds ratio of .988 indicates that students were slightly more likely to have typical or high growth with a younger math teacher than an older math teacher. Specifically, the probability that a student would have typical or high growth in mathematics decreased by 1% as the age of the math teacher increased.

Table 17

2017-2018 Model 1 Hosmer and Lemeshow Test

Model	Chi-square	df	Sig.*
Language Arts	15.829	8	.045
Mathematics	28.620	8	.000

Note: df = degrees of freedom
*$p > .05$

Table 18

2017-2018 Model 1 Model Summary

Model	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Language Arts	2046.171	.006	.008
Mathematics	1967.852	.011	.015

Table 19

2017-2018 Variables in the Equation: *Impact of Teacher Characteristics on Student Growth*

Model	N	B	S.E.	Wald	df	Sig.	Exp(B)	95% Confidence Interval Lower	95% Confidence Interval Upper
LAL Teacher Gender	1,484	-.211	.171	1.512	1	.219	.810	.579	1.133
LAL Teacher Black Ethnicity	1,484	-.254	.112	5.166	1	.023	.776	.624	.966
LAL Teacher Hispanic Ethnicity	1,484	.030	.291	.010	1	.919	1.030	.583	1.820
LAL Teacher Age	1,484	.004	.005	.516	1	.473	1.004	.993	1.014
Math Teacher Gender	1,438	-.157	.151	1.085	1	.298	.855	.636	1.148
Math Teacher Black Ethnicity	1,438	.171	.110	2.413	1	.120	1.187	.956	1.473
Math Teacher Hispanic Ethnicity	1,438	1.260	.371	11.511	1	.001	3.524	1.702	7.295
Math Teacher Age	1,438	-.013	.006	.081	1	.052	.988	.975	1.000

*p < .05

Findings Model 2

Research Question 2 asked: To what extent do the following school-level variables influence student growth: school performance status (Comprehensive schools, Target schools, and NonStatus schools) and percent of student subgroup ethnic composition? In this model we tested how well school-level variables such as school performance statuses and percent of student subgroup compositions predicted student growth in both language arts and mathematics.

Logistic regression models were run for both the 2014-2015 and 2017-2018 school years to determine the extent to which the predictor variables School Comprehensive Status, School Target Status, and School Ethnic Enrollment Subgroups can predict the odds of typical or high growth in the outcome variables LAL SGP and Math SGP. To facilitate understanding of the

findings, the dummy codes for school performance status and school ethnic subgroup are listed below.

1. *School performance status:*

   0 = Comprehensive, 1 = Target and NonStatus;

   0 = Target, 1 = Comprehensive and NonStatus (only in 2017-2018 school year)

2. *Ethnic subgroup:*

   Black enrollment percentage 0 = 69% and less than, 1 = 70% and greater than;

   Hispanic enrollment percentage 0 = 69% and less than, 1 = 70% and greater than;

   Asian/White enrollment percentage 0 = 2% and less than, 1 = 3% and greater than;

Findings 2014-2015

During the 2014-2015 school year, the findings in Table 20 indicate that the school characteristics in the study have a high "fitness" level in both language arts and mathematics according to the Hosmer and Lemeshow Test (Chi Square = .000, df = 2, $p < 1.000$ for both language arts and mathematics). Although the model passed for its "fitness" in language arts, Table 21 presents the model summary statistics, which indicated that the school characteristic predictors can explain only between 0.4% and 0.6% of the variance in the outcome variable LAL SGP. However, the mathematics model summary statistics indicated that the school characteristic predictors can explain between 2.4% and 3.2% of the variance in the outcome variable Math SGP.

In the logistic regression model for predicting typical or high growth in mathematics, only school ethnicity subgroups had a statistically significant relationship to growth. Table 22 indicates that with respect to the percentage of Black student enrollment in schools in the mathematics model, the odds ratio of .555 was significant (Wald = 12.239, p < .000, [CI = .399 – .772]). The odds

ratio of .555 indicates that students registered in a school with a Black student enrollment under 70% were less likely to have typical or high growth in mathematics than students registered in a school with a Black student enrollment greater than or equal to 70%. Specifically, the probability that a student would have typical or high growth in mathematics decreased by 45% when that student was registered in a school with a Black student enrollment under 70%. In addition, with respect to the percentage of Hispanic student enrollment in schools in the mathematics model, the odds ratio of 1.672 was significant (Wald = 5.966, $p < .015$, [CI = 1.107 – 2.524]). The odds ratio of 1.672 indicates that students registered in a school with a Hispanic student enrollment under 70% were more likely to have typical or high growth in mathematics than students registered in a school with a Hispanic student enrollment greater than or equal to 70%. Specifically, the probability that a student would have typical or high growth in mathematics increased by 67% when that student was registered in a school with a Hispanic student enrollment below 70%. In language arts, none of the school demographic predictors were found to be significant.

Table 20

2014-2015 Model 2 Hosmer and Lemeshow Test

Model	Chi-square	df	Sig.*
Language Arts	.000	2	1.000
Mathematics	.000	2	1.000

Note: df = degrees of freedom
*$p > .05$

Table 21

2014-2015 Model 2 Model Summary

Model	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Language Arts	1560.492	.004	.006
Mathematics	1463.932	.024	.032

Table 22

2014-2015 Variables in the Equation: *Impact of Teacher Characteristics on Student Growth*

							95% Confidence Interval	
Model (N = 1161)	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
LAL School Black Enrollment by Racial Subgroup	-.173	.158	1.203	1	.273	.841	.618	1.146
LAL School Hispanic Enrollment by Racial Subgroup	.263	.208	1.604	1	.205	1.301	.866	1.953
LAL School Asian/White Enrollment by Racial Subgroup	.084	.149	.319	1	.572	1.087	.813	1.455
Math School Black Enrollment by Racial Subgroup	-.589	.168	12.239	1	.000	.555	.399	.772
Math School Hispanic Enrollment by Racial Subgroup	.514	.210	5.966	1	.015	1.672	1.107	2.524
Math School Asian/White Enrollment by Racial Subgroup	-.172	.151	1.287	1	.257	.842	.626	1.133

*p < .05

Findings 2017-2018

During the 2017-2018 school year, the findings in Table 23 indicate that the school characteristics in the study have a high "fitness" level in both language arts and mathematics according to the Hosmer and Lemeshow Test (Chi Square = 3.008, df = 7, $p < .884$ for language arts and Chi Square = 11.935, df = 6, $p < .063$ for mathematics). Although the language arts and mathematics models passed for "fitness," Table 24 presents the model summary statistics, which

indicated that the school characteristic predictors can explain only between 0.5% and 0.7% of the variance in the outcome variable LAL SGP; the school characteristic predictors can explain only between 0.2% and 0.3% of the variance in the outcome variable Math SGP.

In the logistic regression model for predicting typical or high growth in both language arts and mathematics, it was found that there were no school demographic variables that had a statistically significant relationship to growth. Based on these findings, the school performance status and school ethnic subgroup characteristics in the study should not be used to explain a student's growth in language arts or mathematics.

Table 23

2017-2018 Model 2 Hosmer and Lemeshow Test

Model	Chi-square	df	Sig.*
Language Arts	3.008	7	.884
Mathematics	11.935	6	.063

Note: df = degrees of freedom
*$p > .05$

Table 24

2017-2018 Model 2 Model Summary

Model	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Language Arts	2048.932	.005	.007
Mathematics	2034.915	.002	.003

Findings Model 3

Research Question 3 asked: How is student growth in language arts and mathematics impacted by a teacher's effectiveness as measured by the practice score received when one controls for teacher- and school-level characteristics? In this model we tested how well the main

predictor variable (teacher practice) predicted student growth while controlling for the previously tested teacher- and school-level variables in both language arts and mathematics.

To test for the main predictor variable's explanatory power when controlling for the variables from the previous two models, a hierarchical logistic regression model was run for both the 2014-2015 and 2017-2018 school years to determine how much value-added the main predictor variable teacher practice score had on the outcome variables LAL SGP and Math SGP. To facilitate understanding of the findings, the dummy codes for the variables in previous models are listed below. Block 1 only added teacher characteristics, Block 2 added school characteristics, and Block 3 added the main predictor variable teacher practice score.

1. *Teacher Characteristics:* gender: 0 = male 1 = female; ethnic background: 0 = Black 1 = all others; 0 = Hispanic 1 = all others; 0 = White/Asian 1 = all others; age: Scale
2. *School Characteristics:* School performance status: 0 = Comprehensive, 1 = Target and NonStatus; 0 = Target, 1 = Comprehensive and NonStatus (only in 2017-2018 school year); Ethnic subgroup: Black enrollment percentage 0 = 69% and less than, 1 = 70% and greater than; Hispanic enrollment percentage 0 = 69% and less than, 1 = 70% and greater than; Asian/White enrollment percentage 0 = 2% and less than, 1 = 3% and greater than
3. *Teacher Practice:* Scale variable

Findings 2014-2015

During the 2014-2015 school year, the findings in Table 25 indicate that adding the school characteristics and then teacher practice in the language arts model increases the "fitness" level according to the Hosmer and Lemeshow Test (Chi Square = 8.052, df = 8, $p < .428$). In mathematics, however, the third Block had a low "fitness" level according to the Hosmer and Lemeshow Test (Chi Square = 19.794, df = 7, $p < .006$).

In the logistic regression model for predicting typical or high growth in language arts, Table 26 presents the model summary statistics, which indicated that there was no change in how much teacher practice can predict the outcome variable LAL SGP. Specifically, the teacher's practice score can explain only between 1.1% and 1.4% of the variance in the outcome variable LAL SGP. However, the mathematics model summary statistics indicated that there was an approximate 1% increase in the explanation of how much teacher practice can predict the outcome variable Math SGP. Specifically, the teacher's practice score can explain between 3.7% and 4.9% of the variance in the outcome variable Math SGP.

Consequently, in language arts, LAL Teacher Practice Score was not a significant predictor variable and did not help explain the variance in the outcome variable LAL SGP. Table 27, on the other hand, indicates that with respect to the Math Teacher Practice Score variable in the mathematics model, the odds ratio of 2.150 was significant (Wald = 6.838, $p < .009$, [CI = 1.211 – 3.816]). The odds ratio of 2.150 indicates that students who had a teacher with a high Math Teacher Practice Score were more likely to have typical or high growth in mathematics than students who had a teacher with a low Math Teacher Practice Score. Specifically, the probability that a student would have typical or high growth in mathematics increased by 215% when that student had a teacher with a high Math Teacher Practice Score.

Table 25

2014-2015 Model 3 Hosmer and Lemeshow Test

Model	Chi-square	df	Sig.*
Language Arts Block 1	21.512	7	.003
Language Arts Block 2	8.664	8	.371
Language Arts Block 3	8.052	8	.428
Mathematics Block 1	58.741	8	.000
Mathematics Block 2	46.954	8	.000
Mathematics Block 3	19.794	7	.006

Note: df = degrees of freedom
*$p > .05$

Table 26

2014-2015 Model 3 Model Summary

Model	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Language Arts Block 1	1013.421	.009	.012
Language Arts Block 2	1012.322	.011	.014
Language Arts Block 3	1012.242	.011	.014
Mathematics Block 1	1082.527	.004	.005
Mathematics Block 2	1062.852	.029	.038
Mathematics Block 3	1055.919	.037	.049

Table 27

2014-2015 Variables in the Equation: *Impact of Teacher Characteristics on Student Growth*

Model: *Block 3* (N = 784)	B	S.E.	Wald	df	Sig.	Exp(B)	95% Confidence Interval Lower	Upper
Math Teacher Gender	.444	.199	4.959	1	.026	1.558	1.055	2.303
Math Teacher Black Ethnicity	-.306	.158	3.751	1	.053	.736	.540	1.004
Math Teacher Age	-.009	.008	1.018	1	.313	.991	.975	1.008
Math School Black Enrollment by Racial Subgroup	-.375	.223	2.846	1	.092	.687	.444	1.063
Math School Hispanic Enrollment by Racial Subgroup	.786	.245	10.259	1	.001	2.195	1.357	3.550
Math School Asian/White Enrollment by Racial Subgroup	.037	.182	.041	1	.840	1.037	.727	1.481
Math Teacher Practice Score	.766	.293	6.838	1	.009	2.150	1.211	3.816

*p < .05

Findings 2017-2018

During the 2017-2018 school year, the findings in Table 28 indicate that adding the school characteristics and then teacher practice in the language arts model increases the "fitness" level according to the Hosmer and Lemeshow Test (Chi Square = 12.736, df = 8, $p < .121$). In mathematics, however, the third Block had a low "fitness" level according to the Hosmer and Lemeshow Test (Chi Square = 49.659, df = 8, $p < .000$).

Table 29 presents the model summary statistics, which indicated that in both language arts and mathematics, there was an approximate 3.3% increase in the explanation of the variance in the outcome variables LAL SGP and Math SGP after adding the main predictor variable, teacher practice, to the overall model. Specifically, in language arts, the teacher's practice score can explain between 3.4% and 4.6% of the variance in the outcome variable LAL SGP; in

mathematics, the teacher's practice score can explain between 4.0% and 5.3% of the variance in the outcome variable Math SGP.

In addition, Table 30 and Table 31 indicate that both the LAL and math teacher's practice score, respectively, were significant influences on predicting student growth. With respect to the LAL Teacher Practice Score variable in the language arts model, the odds ratio of 3.516 was significant (Wald = 34.250, $p < .000$, [CI = 2.307 – 5.356]). The odds ratio of 3.516 indicates that students who had a teacher with a high LAL Teacher Practice Score were more likely to have typical or high growth in language arts than students who had a teacher with a low LAL Teacher Practice Score. Specifically, the probability that a student would have typical or high growth in language arts increased by 352% when that student had a teacher with a high LAL Teacher Practice Score. With respect to the Math Teacher Practice Score variable in the mathematics model, the odds ratio of 4.071 was significant (Wald = 34.627, $p < .000$, [CI = 2.551 – 6.499]). The odds ratio of 4.071 indicates that students who had a teacher with a high Math Teacher Practice Score were more likely to have typical or high growth in mathematics than students who had a teacher with a low Math Teacher Practice Score. Specifically, the probability that a student would have typical or high growth in mathematics increased by 407% when that student had a teacher with a high Math Teacher Practice Score.

Table 28

2017-2018 Model 3 Hosmer and Lemeshow Test

Model	Chi-square	df	Sig.*
Language Arts Block 1	15.829	8	.045
Language Arts Block 2	13.006	8	.112
Language Arts Block 3	12.736	8	.121

Mathematics Block 1	28.620	8	.000
Mathematics Block 2	36.526	8	.000
Mathematics Block 3	49.659	8	.000

Note: df = degrees of freedom
*$p > .05$

Table 29

2017-2018 Model 3 Model Summary

Model	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Language Arts Block 1	2046.171	.006	.008
Language Arts Block 2	2040.113	.010	.013
Language Arts Block 3	2003.126	.034	.046
Mathematics Block 1	1967.852	.011	.015
Mathematics Block 2	1962.633	.015	.020
Mathematics Block 3	1925.430	.040	.053

Table 30

2017-2018 Variables in the Equation: *Impact of Teacher Characteristics on LAL Student Growth*

							95% Confidence Interval	
LAL Model: *Block 3* (N = 1483)	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Teacher Gender	-.428	.187	5.256	1	.022	.651	.452	.940
Teacher Black Ethnicity	-.133	.116	1.302	1	.254	.876	.697	1.100
Teacher Hispanic Ethnicity	.062	.296	.044	1	.835	1.064	.595	1.901
Teacher Age	.004	.005	.613	1	.434	1.004	.994	1.015
School Comprehensive Status	-.204	.137	2.220	1	.136	.815	.624	1.066
School Target Status	.158	.174	.824	1	.364	1.171	.833	1.646
School Black Enrollment by Racial Subgroup	-.297	.148	4.037	1	.045	.743	.556	.993
School Hispanic Enrollment by Racial Subgroup	-.194	.143	1.835	1	.176	.823	.622	1.091
School Asian/White Enrollment by Racial Subgroup	.039	.140	.078	1	.781	1.040	.790	1.368
Teacher Practice Score	1.257	.215	34.250	1	.000	3.516	2.307	5.356

*$p < .05$

Table 31

2017-2018 Variables in the Equation: *Impact of Teacher Characteristics on Math Student Growth*

							95% Confidence Interval	
Math Model: *Block 3* (N = 1433)	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Teacher Gender	-.286	.159	3.246	1	.072	.751	.550	1.025
Teacher Black Ethnicity	.118	.119	.979	1	.322	1.125	.891	1.421
Teacher Hispanic Ethnicity	1.015	.378	7.198	1	.007	2.759	1.314	5.789
Teacher Age	-.007	.007	1.074	1	.300	.993	.980	1.006
School Comprehensive Status	-.173	.153	1.273	1	.259	.841	.623	1.136
School Target Status	-.071	.189	.140	1	.708	.932	.643	1.349
School Black Enrollment by Racial Subgroup	-.123	.153	.643	1	.422	.885	.656	1.194
School Hispanic Enrollment by Racial Subgroup	.029	.142	.041	1	.840	1.029	.779	1.361
School Asian/White Enrollment by Racial Subgroup	.011	.145	.005	1	.942	1.011	.760	1.343
Teacher Practice Score	1.404	.239	34.627	1	.000	4.071	2.551	6.499

*p < .05

Findings Question 4

Research Question 4 asked: To what extent does the relationship between teacher effectiveness and student outcome in language arts and mathematics vary from year 2 of AchieveNJ to year 5 of AchieveNJ? Question 4 analyzed the findings from 2014-2015 (Year 2 of Achieve NJ) and those from 2017-2018 (Year 5 of Achieve NJ) in both language arts and mathematics and identified any significant changes in the relationship between teacher practice scores and student growth. Although most of the students in each school year were different, approximately 56 teachers overlapped in the findings from both of the school years.

In language arts, the main predictor variable of teacher practice was not found to be significant during the 2014-2015 school year, as displayed in Table 32, but during the 2017-2018 school year with respect to the LAL Teacher Practice Score variable, the odds ratio of 3.516 was

significant (Wald = 34.250, p < .000, [CI = 2.307 – 5.356]). Essentially, not only was the LAL Teacher Practice Score not significant during the 2014-2015 school year, but the probability that a student would have typical or high growth in language arts increased by 352% when that student had a teacher with a high LAL Teacher Practice Score during the 2017-2018 school year.

In mathematics, the Math Teacher Practice Score variable during both school years was found to be significant as displayed in Table 32. In addition, there was a significant increase in how much the main predictor variable could explain the outcome variable of Math SGP. Students who had a teacher with a high Math Teacher Practice Score were more likely to have typical or high growth in mathematics in the 2017-2018 school year than during the 2014-2015 school year. Specifically, the probability that a student who had a math teacher with a high teacher practice score would have typical or high growth in mathematics increased by 192% from the 2014-2015 school year to the 2017-2018 school year.

Table 32

Logistic Regression of the Impact of Teacher Practice on Student Growth

Model: *Block 3*	Wald	Sig.	Odds
2014-2015 LAL Teacher Practice Score	.080	.777	1.094
2017-2018 LAL Teacher Practice Score	34.250	.000	3.516
2014-2015 Math Teacher Practice Score	6.838	.009	2.150
2017-2018 Math Teacher Practice Score	34.627	.000	4.071

*p < .05

One means of assessing how well the logistic model performs is to compare the predictions that the model made to observed outcomes in the data. The classification tables for the samples of students' growth level in language arts and mathematics during the 2014-2015 and 2017-2018 school years displayed in Table 33 and Table 34 provide a measure of the discriminative efficiency of the logistic regression model.

Table 33 indicates that the logistic regression model marginally classifies the majority of students in both samples. Roughly 16.6% (122) of those in the 2014-2015 school year who are predicted to have low growth in language arts (296) actually persist and demonstrate typical or high growth, but 27% (199) of those predicted to have typical or high growth in language arts (440) fail to do so. In the 2017-2018 school year, these figures are 16.7% (247) and 26% (385), respectively. Overall, the logistic model correctly classifies 56.4% of students' growth levels in the 2014-2015 school year and 57.4% in the 2017-2018 school year.

Table 34 indicates that the logistic regression model marginally classifies the majority of students in both samples. Roughly 23.2% (182) of those in the 2014-2015 school year who are predicted to have low growth in mathematics (460) actually persist and demonstrate typical or high growth, but 16.6% (130) of those predicted to have typical or high growth in mathematics (324) fail to do so. In the 2017-2018 school year, these figures are 18.4% (264) and 23.5% (337), respectively. Overall, the logistic model correctly classifies 60.2% of students' growth levels in the 2014-2015 school year and 58.1% in the 2017-2018 school year.

Table 33

Classification Table of Predicted vs. Observed Outcomes of the Logistic Regression Model for Student Growth in Language Arts

	Predicted Outcome			
Observed Outcome	Low Growth	Typical/High Growth	N	Percentage Correct
2014-2015 LAL SGP Growth Level				
Low Growth	174	199	373	46.6
Typical/High Growth	122	241	363	66.4
Overall			736	56.4
2017-2018 LAL SGP Growth Level				
Low Growth	339	385	724	46.8
Typical/High Growth	247	512	759	67.5
Overall			1483	57.4

Table 34

Classification Table of Predicted vs. Observed Outcomes of the Logistic Regression Model for Student Growth in Mathematics

Observed Outcome	Predicted Outcome: Low Growth	Predicted Outcome: Typical/High Growth	N	Percentage Correct
2014-2015 Math SGP Growth Level				
Low Growth	278	130	408	68.1
Typical/High Growth	182	194	376	51.6
Overall			784	60.2
2017-2018 Math SGP Growth Level				
Low Growth	348	337	685	50.8
Typical/High Growth	264	484	748	64.7
Overall			1433	58.1

Summary of Quantitative Findings

Presented in this chapter were the findings for three models that examined the relationship between student academic growth in language arts and mathematics while controlling for teacher practice and teacher and school-level characteristics in 2014-2105 and 2017-2018. The findings suggest that very few teacher and school-level variables are significant predictors of student growth.

In the 2014-2015 school year, LAL Teacher Gender ($p < .018$) was the only significant teacher variable in either of the teacher characteristic models for language arts and mathematics. In the 2017-2018 school year, LAL Teacher Black Ethnicity ($p < .023$) and Math Teacher Hispanic Ethnicity ($p < .001$) were the only significant variables in either of the teacher characteristic models for language arts and mathematics.

In the 2014-2015 school year, Math School Black Enrollment by Racial Subgroup ($p < .000$) and Math School Hispanic Enrollment by Racial Subgroup ($p < .015$) were the only

significant school-level variables in either of the school characteristic models for language arts and mathematics. In the 2017-2018 school year, there were no significant school-level variables found in either of the school characteristic models for language arts and mathematics.

In both language arts and mathematics, a teacher's practice score was considerably more significant in the 2017-2018 school year compared to 2014-2015 school year. In addition, in language arts, the probability that a student during the 2017-2018 school year would have typical or high growth in language arts increased by 352% when that student had a teacher with a high LAL Teacher Practice Score. Conversely, LAL Teacher Practice Score was not significant during the 2014-2015 school year and added no value to the explanation of the variance in the overall model. In mathematics, however, both school years found Math Teacher Practice Score to be a significant variable and can help explain the outcome variable of Math SGP. Specifically, in the 2017-2018 school year, the probability that a student would have typical or high growth in mathematics increased by 407% when that student had a teacher with a high Math Teacher Practice Score. Furthermore, students who had a teacher with a high Math Teacher Practice Score were more likely to have typical or high growth in mathematics during the 2017-2018 school year than during the 2014-2015 school year. Specifically, the probability that a student who had a math teacher with a high teacher practice score would have typical or high growth in mathematics increased by 192% from the 2014-2015 school year to the 2017-2018 school year.

CHAPTER V

SUMMARY, RECOMMENDATIONS, AND CONCLUSIONS

Chapter V includes a synopsis of the concept that links teacher effectiveness with student achievement within the research literature, a summary of the research findings, recommendations for further research, and a conclusion.

The notion that teacher effectiveness can be linked positively to student growth measures has been supported in research, such as a study that was conducted through Cincinnati's Teacher Evaluation System (TES) to evaluate the effectiveness of the Danielson Framework to demonstrate a relationship to student achievement. Kane et al. (2011) determined that classroom observations can capture elements of teaching that are related to student achievement. Their findings demonstrated a positive and non-trivial relationship between TES scores and student achievement growth. Their main results indicated that moving from an overall TES rating of "Basic" to "Proficient" or from "Proficient" to "Distinguished" is associated with student achievement gains of about one sixth to one fifth of a standard deviation. For example, if a student started the year at the 50th percentile in math and reading and had a teacher in the lowest quartile of the overall TES rating, and a similar student had a teacher in the upper quartile of that rating, we would expect the second student to be four percentile points ahead in math and five percentile points ahead in reading by the end of the year.

As states move forward with establishing teacher evaluation systems, VAMs have become a key component for most of these systems to measure accurate teacher effectiveness. VAMs attempt to predict the "value" a teacher adds to his or her students' learning growth that standardized assessments have measured. Some states have mandated that up to 50% of the teacher evaluation be tied to student test scores using a value-added measure. The logic of using

teacher evaluation to measure teacher effectiveness for school improvement is based on the positive relationship between teacher quality and student academic growth. Administrators collect data on teacher classroom behavior through classroom observations and compare the results against teacher practice standards on an identified teacher evaluation rubric. Student growth models are then used to measure the amount of growth a student shows from the previous year. This information determines retention, promotion, compensation, and tenure. The use of VAMs for these high-stakes consequential decisions has caused many to question its reliability, validity, and consistency.

Milanowski (2004) examined the teacher evaluation system in Cincinnati to determine the relationship between the evaluation scores of teachers and VAMs of student learning in grades 3–8. The school system's administrators "want[ed] to be justified in inferring that teachers with high scores [were] better performers, defined as producing more student learning" (p. 39). The study yielded some positive and mixed results. However, Milanowski determined that the "moderate level of criterion-related validity" (p. 49) was adequate to support the use of student achievement data in the evaluation of teachers.

Additional studies, however, found little significance when they examined the relationship between student achievement and teacher evaluation ratings. White's (2004) study in Coventry, Rhode Island sought to "describe the relationship between a teacher's overall evaluation score and his or her students' achievement, while controlling for prior achievement, in order to determine the criterion-related validity of the evaluation scores" (p. 3). The researcher analyzed the value-added achievement data in reading and math from 3,617 students and evaluation data for 173 teachers in four elementary school grades for 2 school years. White's results "indicated a small overall correlation in reading (0.240) and essentially no correlation in

math (0.032). The results also indicated rather large fluctuations in correlations between years and across subjects and grade levels" (p. 6). Again, the overall pattern of results provided weak empirical evidence to support the relationship between student achievement and teacher evaluation in elementary schools. Hallinger et al. (2014) concluded that the ideology of using VAMs was stronger than the actual evidence of their impact.

In 2012, New Jersey passed the TEACHNJ Act, which had a goal to "raise student achievement by improving instruction through the adoption of evaluations that provide specific feedback to educators, inform the provision of aligned professional development, and inform personnel decisions" (TEACHNJ Guide, p. 1). These new evaluation policies were a major shift from prior evaluation procedures. The policies eliminated the binary method of evaluation (i.e., satisfactory vs. unsatisfactory), which did not take student growth into account. The new system, however, now incorporates student growth along with a teacher's practice to produce an overall performance score that reflects one of four different teacher rating categories (highly effective, effective, partially effective, and ineffective).

In addition to the teacher's practice evaluation score, student achievement measures are calculated and incorporated into a teacher's summative evaluation. One of the student achievement measures included in a teacher's evaluation consisted of student growth percentiles (SGPs), which were used as the dependent variable in this study. SGPs measure student achievement gains in grades 4–8 in language arts and in grades 4–7 in mathematics (tested grades and subjects) on the Partnership for Assessment of Readiness for College and Careers (PARCC) state assessment, which was adopted and administered to all students in New Jersey in the spring of 2014. An SGP score is a number on a scale from 1 to 99 that measures the change

in a student's achievement from any given year to the next compared to all other students in the state, or "academic peers," who had similar historical results.

As the federal government and state administrations remain focused on student achievement on state assessments and teacher evaluations as a measure of teacher effectiveness, the public policy debate will strengthen around the validity of using value-added measures for tenure, retention, and termination. More specifically, debates will focus on whether teachers should be evaluated based on student achievement, especially when some researchers are skeptical about using value-added measures. This skepticism persists because it has been shown that teacher performance can fluctuate over time depending on several factors that influence student growth such as attendance rate, mobility rate, curriculum material, instructional time, class size, and home and community supports (Darling-Hammond et al., 2012).

The purpose of this study was to examine the relationship between teacher practice and student growth while controlling for teacher- and school-level characteristics. The study determined the probability that students will receive a typical or high growth SGP score on the state assessment in relation to their teacher's practice score based on the practice portion of the evaluation instrument. The study determined the value that teacher practice, teacher characteristics, and school characteristics added to student achievement in the content areas of language arts and mathematics in grades 4–7.

This study is unique in that using scores of students with like scores across the state of New Jersey measured student academic growth. Students are compared to their "academic peers" to determine growth regardless of their level of proficiency, their socioeconomic background, and whether student assignments to teachers are randomized. Furthermore, this methodology allowed the determination of how students of the same academic peer group either grow faster

("high growth") or make less progress ("low growth"). The change in student growth was reported as an SGP on a scale from 1 to 99. A student's SGP growth can be categorized as low (SGP < 35), typical (SGP > 34 and SGP < 66), or high (SGP > 65; New Jersey Department of Education, 2015). Further, the study explained the relationship between teacher practice and student growth in an urban school district that the state of New Jersey identified as a district in need of improvement.

The research literature has also noted a wide range of issues related to the validity and reliability of using value-added measures for the purposes of teacher evaluation. Darling-Hammond et al.'s study (2012) found three key limitations when using VAMs in teacher evaluations:

1. Value-added models of teacher effectiveness are inconsistent.
2. Teachers' value-added performance is affected by the students assigned to them, thereby calling into question the transparency and fairness of using value-added measures of student learning in evaluations.
3. Value-added ratings cannot disentangle the many influences on student progress and thereby provide an incomplete and inaccurate measure of an individual teacher's effectiveness.

The first three models tested in this study controlled for teacher- and school-level characteristics. The findings in Model 1 and Model 2 suggest that very few teacher and school-level variables are significant predictors of student growth, respectively. For Model 1 in the 2014-2015 school year, LAL Teacher Gender ($p < .018$) was the only significant teacher variable in either of the teacher characteristic models for language arts and mathematics. In the 2017-2018 school year, LAL Teacher Black Ethnicity ($p < .023$) and Math Teacher Hispanic Ethnicity

($p < .001$) were the only significant variables in either of the teacher characteristic models for language arts and mathematics. For Model 2 in the 2014-2015 school year, Math School Black Enrollment by Racial Subgroup ($p < .000$) and Math School Hispanic Enrollment by Racial Subgroup ($p < .015$) were the only significant school-level variables in either of the school characteristic models for language arts and mathematics. In the 2017-2018 school year, there were no significant school-level variables found in either of the school characteristic models for language arts and mathematics.

Model 3 was the full model tested on the 2014-2015 and 2017-2018 school years to better interpret the effect that value-added teacher practice had on the explanation of the variance of student growth as measured by SGP when controlling for teacher and school characteristics. In both language arts and mathematics, a teacher's practice score was considerably more significant in the 2017-2018 school year compared to the 2014-2015 school year. In addition, in language arts, the probability that a student during the 2017-2018 school year would have typical or high growth in language arts increased by 352% when that student had a teacher with a high LAL Teacher Practice Score. Conversely, the LAL Teacher Practice Score was not significant during the 2014-2015 school year and added no value to the explanation of the variance in the overall model. In mathematics, however, both school years found the Math Teacher Practice Score to be significant and of help to explain the outcome variable Math SGP. Specifically, in the 2017-2018 school year, the probability that a student would have typical or high growth in mathematics increased by 407% when that student had a teacher with a high Math Teacher Practice Score. Furthermore, students who had a teacher with a high Math Teacher Practice Score during the 2017-2018 school year were more likely to have typical or high growth in mathematics than during the 2014-2015 school year. Specifically, the probability that a student who had a math

teacher with a high teacher practice score would have typical or high growth in mathematics increased by 192% from the 2014-2015 school year to the 2017-2018 school year.

It is important to note that many researchers have illustrated how VAMs are sensitive and can fluctuate substantially within schools even when a different model is used or tested (Amrein-Beardsley & Collins, 2012). One study looked at different growth models to calculate the value-added scores for teachers in the Los Angeles Unified School District (LAUSD) and found that 40% to 55% of the teachers would receive different scores with an alternative model (Briggs & Domingue, 2011). This study, however, adds to the body of research and tests the assumptions derived from the existing literature that the result of teacher practice (the central variable of interest in the study) can partially explain student academic growth.

Recommendations for Future Research

The following recommendations are based on the results of this study:

1. A qualitative study within the same district will be valuable to examine teacher perceptions on the impact of teacher practice scores and how they are incorporated into teacher evaluations.
2. A follow-up study in roughly 3 years in the same district should analyze the same interactions between students' academic growth, as measured by SGP, and their teacher's practice score.
3. A similar study should be replicated that includes other low-performing urban districts to determine if similar findings hold true.
4. A study should investigate the impact of teacher practice on student growth using other school-level variables than those that were used in this study. Within this study, there were

no significant findings during the 2017-2018 school year for how school-level characteristics impacted student growth.

Conclusion

Across the United States, there has been a significant increase in education reform, specifically around how student growth models are used to measure a teacher's effectiveness. At this point, almost every state has either begun developing or implementing growth models and VAMs to better determine a teacher's effectiveness. As a result of this study, it would be interesting to learn how much of the data pertaining to growth measures are being used to inform district policy, and whether those figures have an influence on a school leaders' decision-making. Questions to examine are: Do district leaders use the data in a formative way to inform practice and programs? Are the data being used to make staffing decisions such as non-renewals or tenure charging tenured staff? If the data indicate that effective teachers are positively correlated to student growth, what are the influences that improve how a teacher's practice is observed in relation to that teacher's students' academic growth?

Although the body of research in this area has continued to increase and has still not been consistent, one can conclude from the findings in this study that the effectiveness of a teacher as measured through observations plays a significant role in students' academic growth after 5 years of implementing the new Achieve NJ evaluation mandates. Generally, as the effectiveness of a teacher's practice increases, their students have a greater probability of earning either typical or high growth scores on standardized state tests. Inversely, as the effectiveness of a teacher's practice decreases, their students' chance of earning either typical or high growth scores on standardized state tests decreases. Although this conclusion might seem obvious, the empirical evidence delineated in this study supports it.

The debate about how to educate our students in the best way will always be ongoing. However, this study provides insight for educational leaders and policymakers on the positive relationship between teacher practice and student growth. It is recommended that this type of research continue to explore how other variables influence student learning growth based on how teachers deliver instruction. As a result, these educational debates have influenced many educational policies and regulations to guide the process on how to measure teacher effectiveness appropriately. Moving forward, when setting policy, district leaders across the nation must continuously take into consideration the importance of improving teaching in every classroom for the benefit of student learning. A district should always strive to have a great leader in every school and a great teacher in every class.

References

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*. Retrieved from https://epaa.asu.edu/ojs/article/view/1096

Archer, K. J., & Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal*, *6*(1), 97–105.

Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories*. Retrieved from http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf

Borman, G., & Kimball, S. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal*, *106*(1), 3–20.

Briggs, D., & Domingue, B. (2011). *Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District teachers by the Los Angeles Times*. Retrieved from https://nepc.colorado.edu/publication/due-diligence

Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, *116*(1). Retrieved from http://www.nysed.gov/common/nysed/files/beardsleyputtinggrowthandvalueaddedmodel.pdf

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*. *Education Policy for Action Series*. Providence, RI: Annenberg Institute for School Reform at Brown University (NJ1).

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

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

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan*, *93*(6), 8-15.

Editorial Project in Education Research Center. (2015, April 10). Issue A-Z: No Child Left Behind: An overview. *Education Week*. Retrieved March 18, 2017 from http://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html?print=1

Finn, J. D. (2002). Small classes in American schools: Research, practice, and politics. *Phi Delta Kappan*, *83*(7), 551. Retrieved from https://www.classsizematters.org/wp-content/uploads/2012/10/finn_2002.pdf

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *Review of Economics and Statistics*, *89*(1), 134-150. Retrieved from http://www.cedr.us/papers/credentials/2007-Can%20Teacher%20Quality.pdf

Goldhaber, D. D., & Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. *Developments in School Finance, 1996,* 197-210.

Goldhaber, D. D., & Brewer, D. J. (1999). Teacher licensing and student achievement. In M. Kanstoroom and C.E. Finn, Jr. (Eds.), *Better teachers, better schools* (pp. 83–102). Washington, DC: The Thomas B. Fordham Foundation.

Gottfried, M. A. (2014) Chronic absenteeism and its effects on students' academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, *19*(2), 53-75.

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, *26*(1), 5-28.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013, January). *Have We Identified Effective Teachers?* Bill & Melinda Gates Foundation. Retrieved from https://files.eric.ed.gov/fulltext/ED540959.pdf

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587-613.

Kannapel, P. J., Clements, S. K., Hibpshman, T., & Taylor, D. (2005). *Inside the black box of high-performing high-poverty schools*. Lexington, KY: Prichard Committee for Academic Excellence.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education,79*(4), 54–78.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, *79*(4), 33–53.

New Jersey Department of Education. (2012). *Educators' Effectiveness*. New Jersey Statutes Annotated 6A:10-4.4. Retrieved from http://www.state.nj.us/education/code/current/title6a/chap10.pdf

New Jersey Department of Education. (2018, October). *User Guide for the Teacher Median Student Growth Percentile Report*. Retrieved from https://www.state.nj.us/education/AchieveNJ/teacher/percentile/mSGPuserguide.pdf

New Jersey Department of Education. (2017, August). *Every Student Succeeds Act*. Retrieved from https://www.state.nj.us/education/ESSA/plan/plan.pdf

New Jersey Department of Education. (2018, January). *Every Student Succeeds Act (ESSA): Technical Guide to Summative Ratings and the Identification of Schools in Need of Support*. Retrieved from https://www.state.nj.us/education/title1/accountability/progress/17/

New Jersey Educator Effectiveness Task Force. (2011). *Interim Report*. Retrieved from https://www.state.nj.us/education/educators/effectiveness.pdf

NJ SMART Education Data System. (2016). *Using Student Growth Percentiles*. Retrieved from http://digitallearning.pcgus.com/njgrowth2016/story.html

Okpala, C. O., Smith, F., Jones, E., & Ellis, R. (2000). A clear link between school and teacher characteristics, student demographics, and student achievement. *Education*, *120*(3), 487-487.

PARCC Technical Report. (2018). *Partnership for assessment of readiness for dollege and careers, final technical report for 2017 administration*. Retrieved from https://parcc-assessment.org/wp-content/uploads/2018/03/PARCC-2017-Technical-Report-Final-03162018.pdf

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education*, *14*(4), 329-339. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.9747&rep=rep1&type=pdf

Sartain, L., Stoelinga, S. R., & Krone, E. (2010). *Rethinking teacher evaluation: Findings from the first year of the Excellence in Teaching Project in Chicago Public Schools*. Retrieved from https://consortium.uchicago.edu/sites/default/files/publications/Teacher%20Eval%20Final.pdf

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, *62*(4), 339–355.

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next*, *12*(4). Retrieved from http://educationnext.org/can-teacher-evaluation-improve-teaching/

Teacher Effectiveness and Accountability for the Children of New Jersey (TEACHNJ) Act, P.L.2012, CHAPTER 26, 2. (2012) Retrieved from http://www.njleg.state.nj.us/2012/Bills/PL12/26_.pdf

U.S. Department of Education. (2009). *Race to the Top Program Executive Summary*. Retrieved March 18, 2017 from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness.* The New Teacher Project. Retrieved from https://files.eric.ed.gov/fulltext/ED515656.pdf.

White, B. (2004). The relationship between teacher evaluation scores and student achievement: Evidence from Coventry. *Peabody Journal of Education*, *79*(4), 33-53. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.624.8160&rep=rep1&type=pdf

White, B. R., Cowhy, J., Stevens, W. D., & Sporte, S. E. (2012). *Designing and implementing the next generation of teacher evaluation systems*. University of Chicago Consortium on Chicago School Research. Retrieved from https://ccsr.uchicago.edu/sites/default/files/publications/Teacher%20Evaluation%20Policy%20Brief1_0.pdf

Wise, A. E., Darling-Hammond, L., McLaughlin, M., & Bernstein, H. (1984). *Teacher evaluation: A study of effective practices.* Santa Monica, CA: The Rand Corporation.

Appendix A

Agreement Number 2018-002

DATA SHARING AGREEMENT BETWEEN
THE CAMDEN CITY SCHOOL DISTRICT AND Nicholas Pillsbury

1. This Data Sharing Agreement ("Agreement") is made by and between Nicholas Pillsbury, PhD student at Seaton Hall University, and the Camden City School District ("District").

2. The terms of this Agreement shall commence on September 4, 2018 and end on September 3, 2019. This Agreement may be terminated by either party upon 30 days' advance written notice. The District may terminate this Agreement for cause, upon shorter than 30 days' advance written notice.

3. Nicholas Pillsbury shall comply with all applicable federal and state laws and regulations protecting the privacy of individuals, including Family Educational Rights Privacy Act ("FERPA"), 20 U.S.C. §1232g., *et seq.*; New Jersey student records access laws found at *N.J.S.A.* 18A:36-19; *N.J.A.C.* 6A:14-1.3; *N.J.A.C.* 6A:14-2.9; and *N.J.A.C.* 6A:32-7.1, *et seq.*, and any other federal or state laws pertaining to student records. Nothing in this Agreement may be construed to allow either party to maintain, use, disclose or share personally identifiable student information in a manner not allowed by state or federal law or regulation.

4. "Confidential Information" shall include any "personally identifiable student information", as that term is defined 34 C.F.R. § 99.3.

5. "Disclose" or "disclosure" means to permit access to or the release, transfer, or other communication of Confidential or Non-Confidential Information contained in education records by any means, including oral, written, or electronic means, to any party except the party identified as the party that provided or created the record.

6. The District may disclose Confidential and Non-Confidential Information to Nicholas Pillsbury for the purpose of carrying out an institutional service or function as described in paragraph 8. Further disclosure by Nicholas Pillsbury of any Confidential or Non-Confidential Information released to Nicholas Pillsbury by the District is prohibited.

7. Nicholas Pillsbury shall not: (i) disclose any Confidential or Non-Confidential Information to any unauthorized third party; (ii) make any use of Confidential or Non-Confidential Information except to perform its obligations under this Agreement; or (iii) make Confidential or Non-Confidential Information available to any of its employees, officers, or agents except those internal Nicholas Pillsbury employees who have been authorized by Nicholas Pillsbury to use the information as a component of their project assignment(s). The term "unauthorized third party" for purposes of this Agreement does not include employees, officers, or agents of the Camden City School District who are authorized to have access to the Confidential and Non-Confidential Information.

8. Nicholas Pillsbury agrees to use Confidential and Non-Confidential Information provided by the District only for the purpose of delivering the institutional services and functions of examining the relationship between teacher practice and student growth while controlling for teacher- and school-level characteristics.

9. Nicholas Pillsbury understands that the Confidential Information is protected under state and federal law and agrees to immediately notify the District if any of the Confidential or Non-Confidential Information is disclosed, either intentionally or inadvertently.

10. Nicholas Pillsbury agrees to protect Confidential and Non-Confidential Information in such a manner that it will be disclosed only to Nicholas Pillsbury staff whose duties under this Agreement specifically require them to have access to the Confidential or Non-Confidential Information.

11. Nicholas Pillsbury and the District shall identify at least one authorized representative or data custodian from their respective agencies who shall be responsible for processing and responding to data requests from the other party.

12. Upon request of the District, Nicholas Pillsbury shall agree to permit the District to review or shall provide written assurances to the District regarding the use of Confidential and Non-Confidential Data under this Agreement. The purpose of this provision is to ensure that appropriate policies and procedures are in place to protect the Confidential and Non-Confidential Information and that there has been no further disclosure of the Confidential or Non-Confidential Information.

13. Nicholas Pillsbury, with access to the Confidential and Non-Confidential Information must acknowledge that they are aware of and will abide by the provisions of this Agreement. Nicholas Pillsbury agrees to remove any person from performing work who has, or is suspected to have, violated the terms of this Agreement.

14. By disclosing Confidential and Non-Confidential Information to Nicholas Pillsbury, the District is in no way assigning ownership of the Confidential or Non-Confidential Information to Nicholas Pillsbury. Upon the termination of this Agreement for any reason, Nicholas Pillsbury shall immediately return all Confidential and Non-Confidential Information, including all copies, to the District or destroy all Confidential and Non-Confidential Information in its possession, custody, or control unless otherwise agreed to in writing by both parties. Nicholas Pillsbury will provide the District with affidavits to this effect.

15. This Agreement shall be governed by and construed in accordance with the laws of the State of New Jersey.

16. Violation of this Agreement is cause for immediate termination of this Contract and could subject the violator to civil and criminal penalties.

By signing below, the parties represent that they have read the agreement and agree to be legally bound by its terms and conditions:

Nicholas Pillsbury

9/6/18

Authorized Representative (Date)
PhD Student at Seaton Hall University

Camden City School District

9/7/18

Katrina T. McCombs (Date)
Acting State District Superintendent