The Relationship Between Teacher Practice and Student Performance

Aubrey Johnson

Seton Hall University, aubreyjohnson111@aol.com

Follow this and additional works at: https://scholarship.shu.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
https://scholarship.shu.edu/dissertations/2235
THE RELATIONSHIP BETWEEN TEACHER PRACTICE AND STUDENT PERFORMANCE

By
Aubrey Johnson

Submitted in fulfillment of the requirements for the degree of
Doctor of Education
Department of Education, Management, Leadership and Policy

Under the Supervision of Dr. Elaine Walker

January 2017
SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATE STUDIES

APPROVAL FOR SUCCESSFUL DEFENSE

Aubrey Johnson, has successfully defended and made the required modifications to the


DISSERTATION COMMITTEE
(please sign and date beside your name)

Mentor: Dr. Elaine Walker 8/17/16

Committee Member: Dr. Christopher Tienken 8/17/16

Committee Member: Dr. Soundaram Ramaswami 8/17/16

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

The demand for educational accountability to improve student achievement has been the force behind education reform in recent years. On October 6, 2012, the state of New Jersey enacted the TEACHNJ Act, which reformed teacher tenure laws and required teacher tenure to be linked to their evaluation rating. To support the new tenure reform law, the state of New Jersey revamped its teacher evaluation system and developed an evaluation structure known as AchieveNJ, which allows for the use of multiple measures to evaluate teachers. These measures include components of both teacher practice and student achievement which are calculated to determine an overall summative evaluation teacher score and rating. The 2013–2014 school year was the first full year of implementation of the TEACHNJ Act and the first year the state of New Jersey provided student growth percentile (SGP) scores to be included as a calculated component in teachers’ evaluations.

The purpose of this quantitative study was to examine the relationship between teacher practice and student growth. The study sought to explain the relationships between variables that predict student academic growth. Some of the essential questions regarding this research are as follows: Are teacher-level variables such as grade level taught, gender, and ethnic background significant predictors of student growth? To what extent do the following school-level variables influence student growth: school performance status (Priority schools, Focus schools, NonStatus schools) and percent of student subgroup ethnic composition? How is student growth impacted by a teacher’s effectiveness as measured by the practice score received, when one controls for teacher- and school-level characteristics?

The sample population consisted of 149 language arts (n = 149) and 145 mathematics (n = 145) teachers in grades 4–7. Each teacher in the study received a median SGP score (mSGP)
of their class or course roster. The 294 teacher mSGP scores were reflective of 7,220 students who received a language arts SGP score and 7,163 students who receive a math SGP score. The study involved 30 schools with different grade configurations, performance status, and student ethnic composition.

The research was a cross-sectional study in which ordinal and logistic regression methods were used to test the relationships between the dependent variable (student growth) and independent variables (teacher characteristics, school characteristics, teacher practice). The design consists of three separate models used to answer three research questions. An ordinal regression analysis was used to analyze Model 1 (teacher characteristics on student growth) and Model 2 (school characteristics on student growth). Model 3 is the full model in which a logistic regression analysis was used to better interpret the impact of teacher practice and teacher and school characteristics on student growth. Findings from the data indicated a significant correlation between teacher practice and student growth while controlling for teacher and school characteristics.

This study will help state and district leaders evaluate the mandates put in place and will add to the body of research around teacher evaluations, specifically in urban settings where there often are many economically disadvantaged students.
DEDICATION

Most importantly, this work is dedicated to my mom, Josefina Burke Johnson, who I know is smiling down at me with pride and guiding me throughout this process. It is through your pride, strength, leadership, perseverance, and everything you instilled in me that I was able to complete this endeavor.

Second, I dedicate this work to my family. To my wife, Christine Johnson, for your continual support and encouragement during this journey. To my children, Jaelin, Milan, and Chayce, for understanding what it took to accomplish this task and for allowing me to use this experience to teach you to never stop learning.

To Aubrey Johnson Sr. (Dad), Faye, Norma, and family, I express my gratitude for your support, guidance, and wisdom. To my dear friend Maria Santa for your push, wisdom and belief in me.
ACKNOWLEDGEMENTS

I especially acknowledge my mentor, Dr. Elaine Walker, for your guidance, knowledge, commitment, discipline, and calm throughout this lengthy process. You inspired and motivated me and pushed me to complete this work, and for that I am truly grateful. A special thanks to my committee members, Dr. Christopher Tienken and Dr. Soundaram Ramaswami, for your knowledge, input, and direction.

I would like to acknowledge my Seton Hall Cohort XVIII members who began this process with me. Thank you for your help, support, knowledge, spirit, and fun throughout this process. We supported and helped each other and had many days of great banter. You are like an extended family to me and will always be special!

Lastly, I would like to acknowledge Dr. Donnie Evans for your leadership, mentorship, integrity, and for having a vision that included me.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
</tbody>
</table>

**I. INTRODUCTION**

- Statement of the Problem 10
- Purpose of the Study 11
- Research Questions 12
- Significance of the Study 12
- Theoretical Framework 14
- Limitations of the Study 17
- Delimitations of the Study 17
- Definitions of Terms 18

**II. REVIEW OF RELATED LITERATURE**

- Historical Background 21
- Teacher Effectiveness and Student Achievement 22
- Value-Added Models 25
- Reliability of Value-Added Models 28
- Current Concerns regarding Value-Added Models 33
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. New Jersey Teacher Evaluation Framework</td>
<td>6</td>
</tr>
<tr>
<td>2. Grade Configurations of Schools with Grades 4, 5, 6, and/or 7</td>
<td>46</td>
</tr>
<tr>
<td>3. Instrumentation and Variables</td>
<td>53</td>
</tr>
<tr>
<td>4. Teacher Demographics: Subject Taught, Grade Span, Gender, Ethnicity</td>
<td>59</td>
</tr>
<tr>
<td>5. School Demographics: School Performance Designation, Student Ethnicity</td>
<td>60</td>
</tr>
<tr>
<td>6. Teacher Evaluation Rating: Median Student Growth Percentile, Teacher Practice</td>
<td>60</td>
</tr>
<tr>
<td>7. Model 1 Model Fitting Information</td>
<td>63</td>
</tr>
<tr>
<td>8. Model 1 Test of Parallel Lines</td>
<td>63</td>
</tr>
<tr>
<td>9. Model 1 Pseudo R-Square</td>
<td>63</td>
</tr>
<tr>
<td>10. Parameter Estimates for Impact of Teacher Characteristics on Student Growth</td>
<td>64</td>
</tr>
<tr>
<td>11. Model 2 Model Fitting Information</td>
<td>66</td>
</tr>
<tr>
<td>12. Model 2 Test of Parallel Lines</td>
<td>67</td>
</tr>
<tr>
<td>13. Model 2 Pseudo R-Square</td>
<td>68</td>
</tr>
<tr>
<td>14. Model 2 Parameter Estimates</td>
<td>68</td>
</tr>
<tr>
<td>15. Model 3 Model Summary</td>
<td>71</td>
</tr>
<tr>
<td>16. Model 3 Variables in the Equation</td>
<td>72</td>
</tr>
<tr>
<td>17. Variables Level of Significance</td>
<td>74</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.</td>
<td>Factors influencing student growth</td>
</tr>
<tr>
<td>2.</td>
<td>Models testing the impact on student growth</td>
</tr>
<tr>
<td>3.</td>
<td>Impact of teacher characteristics on student growth</td>
</tr>
<tr>
<td>4.</td>
<td>Impact of school characteristics on student growth</td>
</tr>
<tr>
<td>5.</td>
<td>Impact of teacher characteristics, school characteristics and teacher practice on student growth</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Policy Background

On October 6, 2012, the state of New Jersey authorized the TEACHNJ Act that reformed teacher tenure laws and required that teachers’ tenure be tied to summative evaluation ratings. This decision is a result of numerous conversations, over several years, by education reformers who believed in the need for educational accountability in public schools to improve student achievement. In 1965, the federal Elementary and Secondary Act (ESEA) emphasized equal access to education and high standards for academic performance while demanding more accountability. ESEA was reauthorized in 1994 and evolved into the No Child Left Behind Act of 2001 (NCLB, 2002). At the core of the NCLB Act were measures designed to drive improvements in student achievement and measures that would hold states and schools more accountable for student academic progress. The NCLB Act required:

- **Annual Testing:** Annual assessments that are aligned with state standards in reading/language arts, mathematics, and science in grades 3–8 and in grade 11 at the high-school level.
- **Academic Progress:** States were required to ensure that all students reached proficiency levels (100%) in language arts and mathematics on state tests by the end of the 2013–2014 school year.
- **Report Cards:** States and school districts were required to provide annual report cards showing demographic and assessment information including student achievement.
disaggregated by subgroups based on gender, ethnicity, special needs, and limited English proficiency.

- Teacher Qualifications: Every teacher in core content areas was required to be “highly qualified” by means of certification and teaching proficiency in the subject matter taught.
- Reading First: This competitive-grant based program helped states and districts to establish reading programs and reading intervention initiatives for children in grades K–3, to ensure that every child would reach grade-level proficiency in reading by the end of grade 3.

After a few years of implementation, many educators and policymakers expressed concerns regarding the mandates set forth by NCLB (Editorial Projects in Education Research Center, 2011). Questions were raised regarding the fairness of the NCLB goals, the school-level targets, and the timeframe that required 100% proficiency for all students by the end of the 2013–2014 school year. It was conjectured that most schools within the country would not meet the goal of 100% proficiency for all students and would therefore be labeled as “failing” schools by 2014. In 2011, U.S. Secretary of Education Arne Duncan campaigned to rewrite the law. He created a waiver option for states that wanted to opt out of some of the NCLB mandates. States that participated in the waiver option had the freedom to set their own student achievement goals and design aggressive interventions for the lowest five percent of failing schools (Priority schools). Furthermore, states were required to identify another ten percent of schools that struggled with achievement gaps among specific subgroups of students and low graduation rates (Focus schools). States would establish performance targets for every school and every student subgroup, and then set ambitious but achievable goals. To improve teacher effectiveness, the federal government required states and school districts to collaborate to:
1. establish clear approaches to measuring individual student growth.

2. design and implement rigorous, transparent, and fair evaluation systems for teachers that differentiate effectiveness using multiple rating categories that take into account data on student growth as a significant factor.

3. conduct annual evaluations of teachers that include timely and constructive feedback and provide teachers with data on student growth for their students, classes, and schools. (U.S. Department of Education, 2009, p. 9)

These waiver options were designed to provide some relief to the NCLB mandates and were less prescriptive than other administrations’ education improvement priorities such as the Federal Race to the Top initiative which offered bold incentives to states willing to spur innovative and systemic reforms to improve teaching and learning in schools. Its main goals were to pursue higher academic standards, improve teacher effectiveness with the use of student achievement data to guide instruction in the classroom, adopt new strategies to help struggling schools, and build data systems to support instruction. This initiative spurred the implementation of a new generation of teacher evaluation models across the states that would promote effective teaching practices to raise student achievement and that would offer professional support to the retention of effective teachers.

In 2011, Tennessee became one of the first states that tied student achievement on state-mandated standardized tests of language arts, mathematics, and science to teacher evaluations. Student achievement data accounted for 50% of a classroom teacher’s summative evaluation, and personnel decisions (e.g., promotion, retention, tenure, compensation) were based on these evaluations (Piro, Wiemers, & Shutt, 2011). During the 2011–2012 school year, Tennessee state assessment scores improved, in aggregate, at a faster rate than any previously measured year.
Administrators cited that teacher evaluation played an important role in improving instruction and resulted in higher student achievement gains (Tennessee Department of Education, 2012).

The Illinois Performance Reform Act (PERA), which was signed into law in 2010, required that every school district adopt a teacher evaluation system that linked teacher observations and student growth. The Evanston School District in Illinois was one of the first districts in the state that used student growth as a measure for its teacher evaluation system. The idea was to bring student performance to the forefront of the public school teacher evaluation conversation and to help teachers understand the relationship between their practices and student growth. District administrators viewed the student growth component as an important accountability measure to guarantee that each student obtained one year’s growth for which their assigned teacher was responsible. By aligning teacher evaluation ratings with student achievement on state and district assessments, administrators believed that the student growth measure would address some of the perceived drawbacks of the NCLB accountability system by focusing on every student rather than “just sub-groups of students or those at the borderline of proficiency” (White, Cowhy, Stevens, & Sporte, 2012, p. 21). Teachers perceived a conflict with the implementation of the new teacher evaluation system because of large fluctuation in student growth scores. They also questioned the reliability and validity of the assessments used to interpret student growth and the measures used to determine adequate student growth.

The state of New Jersey participated in the Race to Top initiative and overhauled its evaluation process for teachers to comply with the required provisions of the grant program. New Jersey authorized the TEACHNJ Act that changed teacher tenure laws. A teacher’s tenure would now be dependent on his or her evaluation score, and lifelong tenure was no longer guaranteed.
The law declared:

The goal of this legislation 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;

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)

Prior to the TEACHNJ Act, New Jersey had the oldest tenure law in the country, dating back to 1909 in which teachers were granted automatic tenure after three years and one day of service with no direct link to student achievement. Teachers hired prior to the enactment of the TEACHNJ Act were not affected and earned tenure automatically after three years and one day in the position. However, for teachers hired after August 6, 2012, the tenure determination was based partly on student achievement over a period of four years. As stated in the law:

In order to achieve tenure pursuant to this subsection, a teacher shall also complete a district mentorship program during the initial year of employment and receive a rating of effective or highly effective in two annual summative evaluations within the first three years of employment after the initial year of employment in which the teacher completes the district mentorship program. (Teacher Effectiveness and Accountability for the Children of New Jersey (TEACHNJ) Act, Chapter 26, 9, 2012)

The TEACHNJ Act mandated statewide implementation of a rigorous teacher evaluation system starting in the 2013–2014 school year. To support the new tenure reform law, the state revamped the teacher evaluation system and developed a support structure known as AchieveNJ which allowed for the use of multiple measures of performance to evaluate teachers. The new measure moved away from a single evaluation that rated mainly teacher practice to an evaluation that
measured a combination of teacher practice and student growth. This new system required four categories for teacher ratings (highly effective, effective, partially effective, and ineffective) based on multiple measures of student learning and growth. Multiple observations became required for all teachers, whereas in the past multiple observations were required for non-tenured teachers. Table 1 provides an overview of the evaluation process prior to and after the implementation of the new law.

Table 1

<table>
<thead>
<tr>
<th>New Jersey Teacher Evaluation Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Evaluation Prior to AchieveNJ</td>
</tr>
<tr>
<td>Past</td>
</tr>
<tr>
<td>Binary measurement with limited ability to differentiate effectiveness and inform growth</td>
</tr>
<tr>
<td>Evaluation based solely on single measure (teacher practice)</td>
</tr>
<tr>
<td>Multiple observations (3) required for non-tenured teachers</td>
</tr>
<tr>
<td>Ongoing calibration and monitoring of observations</td>
</tr>
</tbody>
</table>

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

In New Jersey, school district leaders have the latitude to select from several state-approved teacher practice evaluation instruments in accordance with the TEACHNJ Act. The teacher practice evaluation instrument is used to assess the competencies of a teacher practice by gathering evidence, primarily through classrooms observations. Districts can choose from the following notable teacher practice evaluation instruments: Charlotte Danielson Framework for Teaching, Marzano’s Causal Teacher Evaluation Model, Mid-Continent Research for Education
and Learning (McREL) Teacher Evaluation Standards, Stronge Teacher and Leader Effectiveness Performance System, and Focal Point Teaching Practice Model. School districts could 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 required non-tenured teachers to be observed three times annually. This requirement consisted of two long observations (40 minutes minimum) and one short observation (20 minutes minimum) in the first two years of teaching and one long and two short observations in the third and fourth years of employment. Non-tenured teachers were required to be observed by more than one certified administrator (multiple observers), and it was recommended but not required for tenured teachers to have multiple observers. However, multiple observers were required for teachers placed on a corrective action plan. Teachers were automatically placed on a corrective action plan when they were rated ineffective or partially effective on their summative evaluation. Tenured teachers were required to have a minimum of three short observations each year. As per New Jersey state statute:

> To earn a teacher practice score, a teacher shall receive at least three observations. If a teacher is present for less than 40 percent of the total student school days in an academic year, he or she shall receive at least two observations to earn a teacher practice score.

(Educators Effectiveness, 2012, p. 25)

The law required that teachers be evaluated based on the length of time they were instructing students. Appropriate implementation of the evaluation framework based on measures of teaching practice would allow meaningful feedback to teachers on their instructional practice.

In addition to the teacher’s practice evaluation score, student achievement measures were calculated and incorporated into a teacher’s summative evaluation. Student achievement
measures consisted of student growth percentiles (SGPs) and/or student growth objectives (SGOs). SGPs measured student achievement gains in grades 4–8 in language arts and in grades 4–7 in mathematics (tested grades and subjects) on the New Jersey Assessment of Skills and Knowledge (NJ ASK) or the new 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 was a number on a scale from 1 to 99 that measured 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. A teacher’s evaluation reflected the median student growth percentile (mSGP) of all the students in his or her class. A student below grade level with a low proficiency rate could earn a high SGP score, which means that the student demonstrated more growth than his or her “academic peers” and signals that the teacher’s instructional practice may have assisted the student growth.

TEACHNJ required every teacher to set SGO goals for themselves to measure student learning over the course of the year. SGOs were another measure that was factored into a teacher’s summative evaluation rating. These were goals that teachers set for themselves in the beginning of the school year related to student achievement. SGOs were used for the non-tested grades and subjects and should be aligned to the Common Core State Standards. Once SGO goals were set, teachers planned instruction throughout the year that ensured they taught the required standards and used a quality assessment that accurately and fairly measured student performance. The number of SGOs a teacher must set was determined by the content and grade level taught. Teachers who received an mSGP score in grades 4–8 in language arts and in grades 4–7 in mathematics were required to develop one or two SGOs. Teachers who did not receive SGP scores, including math teachers in grade 8, were required to set two SGOs. It was
recommended that teachers who received an mSGP score with 25 or fewer students set two SGOs with the understanding that an mSGP score would not be provided when the student population dropped below 20 students. This was a safety precaution for school districts with high mobility rates.

During the 2013–2014 school year, a teacher’s overall summative evaluation ratings included the following multiple measures: (a) teacher practice score derived from three observations, (b) SGP scores and/or SGO scores which were weighted and added together to calculate an overall summative evaluation score of 1 (ineffective), 2 (partially effective), 3 (effective), or 4 (highly effective). Teachers in tested grades and subjects received an mSGP score, and the overall summative evaluation rating was calculated by combining the multiple weighted measures of teacher practice (55%), mSGP (30%), and SGO (15%). For teachers in non-tested grades and subjects 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 school district in this study, like other districts in the state, had begun complying with the mandates required by the TEACHNJ Act. For two years, the district was one of several school districts that participated in the state’s Evaluation Pilot Advisory Committee (EPAC) which guided the development of the state’s evaluation policy. During the 2013–2014 school year, the district fully implemented the new teacher and principal evaluation system by building on key learning ideas gleaned from the pilot implementation. The district created an implementation timeline to guarantee that teacher observation practices were aligned with the expectations outlined in AchieveNJ. School leaders aimed to improve the quality of the feedback they provided to their staff by strategically focusing on instructional leadership and effective
instructional practices. Teachers attended numerous workshops on the process and procedures of the new evaluation system and received training specific to their teacher performance (practice) rubric with a focus on pedagogical instructional strategies to meet the needs of diverse learners. The district adopted the Department of Education’s state-approved Focal Point Teaching Practice Model instrument to evaluate teacher practices. The instrument focused on the following seven performance domains:

- preparation for instruction,
- use of data to inform instruction,
- delivery of instruction,
- interventions to meet diverse needs,
- classroom environment,
- leadership, and
- professionalism.

In recognition of the district’s solid implementation of the observation and evaluation system, the New Jersey Department of Education invited the district, along with six others, to partner and share best practices with the state as well as with other districts.

**Statement of the Problem**

The new teacher evaluation mandates have required teachers to adapt to a new model of accountability, which determines teacher effectiveness by establishing a relationship between student achievement and teacher evaluations. Because the policy is so new, no research studies in the state have offered substantial insight to examine the link between teacher practice evaluations and SGPs. Some studies have found inconsistencies between an individual teacher’s rating and student performance (Weisberg, Sexton, Mulhern, & Keeling, 2009) and extreme fluctuations in
teacher evaluation rating from year to year. A study by Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012) examined teacher evaluation data from five school districts and found that 20% to 30% of the teachers who were rated less effective in one year were rated the same the following year. Furthermore, 25% to 45% of the teachers rated less effective moved to the highly effective rating the following year. The same was true for those who were rated highly effective in one year; specifically, only a small minority remained in the highly effective rating the following year. Although the district in this study was recognized by the State Department of Education as a leader in its implementation of the new evaluation system, there has not been an examination of the impact of the TEACHNJ Act and the relationship between a teacher’s practice and student growth.

**Purpose of the Study**

The state of New Jersey mandated the implementation of the new teacher evaluation system in 2012, when the district in this study completed its first full year of implementation of the TEACHNJ and Achieve NJ mandates. This study attempted to explain the relationship between teacher practice and student achievement on the statewide assessment to determine the correlation between students who demonstrated typical or high growth on the state assessment and teachers who were rated effective on the practice portion of the evaluation instrument. The study determined 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 purpose of this study was to examine the relationship between teacher practice and student growth while controlling for teacher- and school-level characteristics. Student growth was measured by analyzing the teacher’s mSGP used for evaluative purposes and used to compare student growth across the state from year to year. Student growth was measured by
comparing the change in their achievement on the state assessment from one year to the next when compared to their academic peers, defined as other students with the same historical state assessment results. The change in student growth was reported by the state as an SGP score on a scale from 1 to 99. A student’s SGP score is categorized as low (SGP < 35), typical (SGP > 34 and SGP < 66), or high (SGP > 65). Currently, there are no relevant studies that correlate SGPs to teachers’ practice evaluation ratings using value-added models.

Research Questions

The following questions guided this study and are derived from the heuristic model depicted in Figure 1:

1. Are teacher-level variables such as grade level taught, gender, and ethnic background significant predictors of student growth?
2. To what extent do the following school-level variables influence student growth: school performance status (Priority schools, Focus schools, NonStatus schools) and percent of student subgroup ethnic composition?
3. How is student growth impacted by a teacher’s effectiveness as measured by the practice score received, when one controls for teacher- and school-level characteristics?

Significance of the Study

The federal Race to the Top initiative offered bold incentives to states willing to spur innovative and systemic reform to improve teaching and learning in schools. This initiative spurred the implementation of a new generation of teacher evaluation models across the country that would promote effective teaching practices to raise student achievement and support the retention of effective teachers. The state of New Jersey participated in the Race to the Top
initiative and overhauled its evaluation process for teachers with a goal to raise student achievement by revamping its teacher evaluation process. 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 second year of implementation of the new teacher evaluation framework (AchieveNJ), and any research finding will benefit future policy decisions. There is no current research that has examined the relationship between SGPs and teacher practice. This study will add to the body of research on the effectiveness of teacher practice on student achievement in an urban setting where there is a large number of poor and low-performing students. This study will present recommendations for policymakers and school leaders on its implementation efforts to support administrators and teachers in their efforts to meet the needs of diverse learners in an urban education setting. Significant time and resources have been allocated both at the district and state levels to support the TEACHNJ Act and the AchieveNJ initiative. Thus, it is important that teachers and administrators believe in the reliability and validity of the process. With the federal government and state administration focused on student achievement on state assessments and teacher evaluations as a measure of teacher effectiveness, public policy debate will intensify around using value-added measures for tenure, retention, promotion, performance pay, and termination. More specifically, debates will center on whether teachers should be evaluated based on student achievement, when researchers are skeptical about using value-added measures and when teacher performance can fluctuate over time depending on several factors that influence student growth such as attendance rate, mobility rate, class size, curriculum material, instructional time, prior teacher schooling, and home and community supports (Darling-Hammond et al., 2012).
This issue is particularly salient when one considers that students are not randomly assigned to teachers. This study should improve upon and add to the previous studies that sought to connect teacher effectiveness to student achievement. This study is unique in that student academic growth is measured by using scores of students with like scores across the state of New Jersey. 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. More so, the study will explain the relationship between teacher practice and student growth in an urban school district that was identified as a district in need of improvement by the state of New Jersey.

Theoretical Framework

The new teacher evaluation systems in many school districts represent a departure from prior approaches to teacher evaluations that were procedural and systemic in nature. Past teacher evaluation models typically used checklists with little observational feedback and rarely included data on student achievement (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1985). Teacher evaluations were not linked to a teacher’s tenure, and personnel decisions were linked to degrees, college credits, and years of experience because school-level actors had little faith in the fairness of most observations of teachers (Podgursky & Springer, 2007). A study conducted by Weisberg et al. (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. The use of student data to assess teachers began to be seriously considered in the late 1990s with the advent of evaluation reforms which sought to provide schools with effective systems that encouraged all teachers to engage in a cycle of continuous improvement (Wright, Horn & Sanders, 1997).
However, attaching student achievement to teacher evaluations has created some controversy regarding the validity of using solely student achievement (Kane, Taylor, Tyler, & Wooten, 2010). Is it fair to hold teachers accountable for student achievement on standardized tests when there are many factors outside of their control?

The policy logic of linking teacher evaluation to student achievement has been based on several assumptions. Some research findings have confirmed a direct relationship between teacher effectiveness and student academic success. Stronge, Ward, and Grant (2011) examined classroom practices of effective versus less effective teachers based on student achievement gains in reading and mathematics scores. The grade 4 end-of-course reading and mathematics tests served as the grade 5 pre-test. The results of the study indicated students’ achievement levels in language arts and mathematics were higher for effective teachers as compared to less-effective teachers by more than 30 percentile points. For reading, the difference in gains was 0.59 standard deviations in one year. Students taught by less-effective teachers could expect to score at the 21st percentile on the state’s reading assessment, whereas students taught by effective teachers could expect to score at approximately the 54th percentile. In mathematics, the difference in gain scores was 0.45 standard deviations. Students in the classrooms with less-effective teachers scored, on average at the 38th percentile, while students with effective teachers’ classrooms scored at the 70th percentile. Given the findings from previous research, it is wise to investigate the contributing factors between high- and low-performing teachers and how might they differ in their instructional practices, use of questioning, and classroom organization and management to determine how these factors affect student achievement.

Findings such as these may justify the use of student achievement data in teachers’ performance evaluations. It is safe to assume that the use of multiple measures to evaluate teachers can
capture the impact of a teacher’s effectiveness on student academic growth in a reliable manner. We can assume that the new approach to teacher evaluation will produce reliable and sustainable improvements in the quality of teaching and learning (Darling-Hammond et al., 2012).

In the present study, it is assumed that several factors are likely to impact student performance and that, to understand the influence of teacher practice on student growth, these factors need to be considered. Moreover, these factors are also likely to impact a teacher’s practice as well. These assumptions, derived from the extant literature, stipulate that student growth is a function of three sets of variables: teacher practice (the central variable of interest in the study), teacher characteristics, and school characteristics (see Figure 1 below). Moreover, it is argued that teacher practice is influenced by both the characteristics of the teacher and the school context in which he or she teaches. Indeed, one could argue that a teacher’s practice mediates the influence of both sets of variables. Chapter II of this dissertation reviews the relevant literature to support these assumptions.

Figure 1. Factors influencing student growth. This figure illustrates the three variables (Teacher Characteristics, School Characteristics and Teacher Practice) that influence student growth.
Limitations of the Study

1. A limitation to the research is that the study was conducted in only one year due to the change in state assessment moving from NJ ASK to PARCC assessments.

2. The study excluded students that were identified as special education and bilingual students due to variables that were not controlled for in this study. Such variables included pull out and push in support, the number of years in the program, and the type of classification.

3. The teacher evaluation framework did not have an inter-rater reliability component, and there was variance in professional development given to observers who provided feedback to teachers.

4. The study did not include best instructional practices that teachers would use to lead to an increase in teacher effectiveness and student achievement.

Delimitations of the Study

1. This study was delimitated to general education teachers in grades 4–7 who received an mSGP score which was calculated and provided by the state.

2. This study focused on teachers who taught students in grades 4–7 in language arts or mathematics.

3. Data collection were confined to state assessment results and teacher practice scores from the school district’s McRel system, which maintain teacher’s observation and evaluation scores. School-level data that included teacher and school characteristics were collected online from New Jersey State school performance reports and data was retrieved from the state’s NJSMART portal, which maintains student and teacher records.
**Definition of Terms**

For clarification, the following terms are defined as they were used throughout this study.

*Accountability*: a school and its teachers are held responsible for the performance of its students.

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

*AchieveNJ*: a state mandate that relied on multiple measures of performance to evaluate teachers. These measures included components of both student achievement and teacher practice. While all New Jersey teachers received an annual summative evaluation rating, the components used to determine these ratings varied depending on the grades and subjects that educators taught.

*Evaluation Instrument*: a teaching practice evaluation instrument selected by a school district from a state-approved lists. The evaluation instrument was a rubric that provided measurements that captured teacher competencies. The scores from the evaluation rubric were components of the teacher’s observation that were included in the summative evaluation rating for the teacher.

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

*Socioeconomic Status (SES)*: an economic and sociological combined measure of a grouping of people with similar work experience and social position in relation to others that are based on income, education, and occupation.
**Student Growth Objectives (SGOs):** long-term academic goals for groups of students set by teachers 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 made by that student’s academic peers within a testing year.

**Summative Evaluation:** consisted of two primary components: teacher practice (measured primarily by classroom observations) and student achievement. Under AchieveNJ, teachers were evaluated based on multiple measures of educator practice and student achievement. Each element of the evaluation resulted in a rating of 1 to 4, which was weighted according to the state formulas. Once the scores for all evaluation measures were finalized, each educator received 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 and means by which a classroom teacher delivers instruction.

**Teacher Practice Score:** the average of three or more observations. Teacher practice scores could be 1 (ineffective), 2 (partially effective), 3 (effective), or 4 (highly effective).

**TEACHNJ Act:** the tenure reform law which reformed the processes of earning and maintaining tenure. Under the act, tenure decisions were based on multiple measures of student achievement and teacher practice as measured by new evaluation procedures. All teachers would 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 would be charged with inefficiency. The law declared that the goal was “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).

Value-Added Modeling (VAM): a method of teacher evaluation that measures the teacher’s 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.

Rate: how effectively or how quickly students’ learning achievement improves. Achievement growth is tracked and calculated to determine a student’s growth rate.
CHAPTER II
LITERATURE REVIEW

Historical Background

The literature review examines several areas regarding the connection between teacher effectiveness and teacher evaluations to student academic performance. This review examines the literature surrounding the characteristics of effective teachers by means of their traits and practices that promote academic growth. This chapter explores the new impetus to incorporate value-added models (VAMs) to determine teacher effectiveness and to hold teachers accountable to 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. The literature review focuses primarily on the relationship between teacher quality, effectiveness, and accountability and student achievement. Review of the literature revealed two groups of research studies: findings that encouraged the use of VAMs to determine teacher effectiveness for evaluation purposes and findings that revealed flaws in using VAMs alone to evaluate teachers’ effectiveness. The literature was extracted from academic journals, doctoral dissertations, review articles, and related books via ProQuest search engines within educational and social science databases. Much of the literature was empirical in nature rather than theoretical. This chapter reviews the professional dialogue and the result of studies concerning the use of student achievement to evaluate teacher effectiveness. The current climate of educational reform and current accountability processes that have been put in place in many school districts require a critical review of its impact on the education process.
Teacher Effectiveness and Student Achievement

The Tennessee Department of Education had the first data-tracking system, Tennessee Value-Added Assessment System (TVAAS), in the country that 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. The study conducted by Nye, Hedges, and Konstantopoulos (2001) and by the Tennessee Department of Education examined the effects of smaller classes on student achievement and confirmed that small class size has an impact on student achievement in grades K–3. 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. Also, 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 three 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 three 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 attending inner city schools than for White students attending 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). Minority students tended to have lower achievement scores than White students before participation in small classes and made larger achievement gains by the end of the year (Nye, Hedges, & Konstantopoulos, 2004). Students in the study returned to large-classroom settings for grades 4–8, and there were some carryover effects for those students who attended small-class settings in the primary grades. Finn (2002) 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. Project STAR did not measure classroom processes, but an array of research has indicated that teacher morale is improved in small classes (Johnston, 1990) and teachers spend more time on direct instruction and less on classroom management when classes are small (Molnar, Smith, & Zahorik, 1999).

Sanders and Rivers (1996) confirmed a direct relationship between teacher quality and student achievement. When grade 3 students were placed with three high-performing teachers each year in a row, on average they scored in the 96th percentile on Tennessee’s statewide assessment in mathematics by the end of grade 5. However, when grade 3 students were placed with low-performing teachers three years in a row, their average score on the Tennessee’s statewide mathematics assessment at the end of grade 5 was in the 44th percentile, a difference of 52 percentile points. Conversely, researchers who have reviewed this study have questioned the validity of the reported findings related to teacher effectiveness on student learning (Kupermintz, 2003). The study controlled only for student pre-test results and did not consider the effects of teacher, classroom, and student variables such as ability or social or ethnic characteristics.
In a similar study in Texas, Jordan, Mendro, and Weerasinghe (1997) found that teacher quality explained the largest portion of the difference in reading and math achievement. The results of this study confirmed the findings of Sanders and Rivers (1996), discussed above. However, this study controlled for student and school variables such as ethnicity, language proficiency, gender, and socioeconomic status. Moreover, this study examined different student populations, grade levels, statistical methods, and analysis. They found a 34-percentile-point difference in reading and a 49-percentile-point difference in mathematics between students who had three consecutive years of highly effective teachers compared to those who had three consecutive years of less-effective teachers in the Dallas, Texas schools.

Stronge et al. (2011) examined the characteristics of effective teachers versus less-effective teachers by examining classroom instructional and management practices. Student learning gains were measured for one year where the grade 4 end-of-course reading and mathematics tests served as the grade 5 pre-tests. 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 that “this translated into more than a 30 percentile difference in achievement based on one years 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 a significant difference in disruptive behavior with classrooms taught by less-effective teachers. There was less time on task due to disruptive
behaviors, and less-effective teachers had three times more disruptions than effective teachers. Additionally, 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**

Almost all states are moving forward with growth and VAMs as a key component of their state teacher evaluation systems. VAMs attempt to predict the “value” a teacher adds to his or her students’ learning growth measured by standardized assessments. 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. Evaluations systematically incorporate data on the achievement of the teacher’s students over the preceding year (Gates Foundation, 2013). 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 conducted by
Taylor and Tyler (2012) in a Cincinnati school district found evidence that suggests that midcareer teachers’ effectiveness improved during the school year and subsequent following school years when VAMs were incorporated in the evaluation process. Students in mathematics performed higher on end-of-year math tests the year value-added measures were in place compared to the previous year’s evaluations. Taylor and Tyler (2012) explained:

> These improvements persist and, in fact, increase in the years after evaluation. We estimate that the average teacher’s students score 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.

Education reformers believe that a teacher’s effectiveness can be measured and used for evaluation purposes when using VAMs by controlling for factors that are outside a teacher’s influence, such as prior test results and socioeconomic status. However, researchers have argued that there needs to be fairness when evaluating teachers, especially when comparing students with different socioeconomic backgrounds and classrooms with different demographic attributes such as class size, ethnicity, and push in and pull out programs that impact achievement.
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 taught by effective teachers, but the differences were not significant. They 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?” (Kimball et al. 2004, 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 examining 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). He 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 supporting 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.

**Reliability of Value-Added Models**

The research literature has highlighted a wide range of issues related to the validity and reliability of VAMs. Across the country, school districts are using value-added measures to make key personnel decisions about retention, dismissal, and compensation of teachers; however, there is a major debate amongst researchers on whether VAMs should be used for those purposes. The most commonly used model has been the Education Value-Added Assessment System (EVAAS) model. This model was first developed as the Tennessee Value-Added Assessment System (TVAAS) and adopted in Tennessee in the 1990s (Collins & Amrein-Beardsley, 2014; Sanders & Horn, 1998). Many view a VAM as a complex algorithm that requires high statistical expertise to develop and interpret the results when used to determine how much teachers contribute to student learning. The American Statistical Association makes the following recommendations regarding the use of VAMs:

- VAMs are generally based on standardized test scores, and do not directly measure potential teacher contributions toward other student outcomes.
• VAMs typically measure correlation, not causation: positive or negative effects attributed to a teacher may be caused by other factors that are not captured in the model.

• Under some conditions, VAM scores and rankings can change substantially when a different model or test is used, and a thorough analysis should be undertaken to evaluate the sensitivity of estimates to different models.

• VAMs should be viewed within the context of quality improvement, which distinguishes aspects of quality that can be attributed to the system from those that can be attributed to individual teachers, teacher preparation programs, or schools. Most VAM studies found that teachers account for about 1% to 14% of the variability in test scores, and that many opportunities for quality improvement are found in the school-level conditions. Ranking teachers by their VAM scores can have unintended consequences that reduce quality. (ASA, 2014, p. 2)

An array of different VAMs have been used across the states. It is possible that a teacher’s VAM score could be different from state to state when the same student data is used. 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. This was verified in a similar study conducted by Briggs and Domingue (2011) wherein an alternative statistical model was used to calculate the value-added scores for teachers in the Los Angeles Unified School District (LAUSD) whose scores were 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 by 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.
Researchers have questioned the reliability of VAMs to compare teachers working in very different socioeconomic communities with very different student populations. Teachers in some schools have little access to high-achieving students from affluent families and communities, and teachers in other schools have similarly little access to low-achieving students from poor families and communities. The VAM may not accurately identify teacher effectiveness across a common scale. Another common issue is that VAMs do not measure the effect students have on their own learning. There are “peer effects” arising from whether students reinforce or discourage one another’s academic efforts in a classroom. There are peer effects when small groups of students work collaboratively or when students work against the common goal of learning in a class. Haertel stated:

These kinds of effects are important, of course, but for value-added modeling, there are two additional kinds of peer effects that may be equally or more important. The first of these has to do with how the members of the class collectively influence the teacher’s pacing of instruction, the level at which explanations are pitched, the amount of reading assigned, and so forth. If the teacher is meeting the students where they are, then the average achievement level in the class is going to influence the amount of content delivered to all of the students over the course of the school year. In the real world of schooling, students are sorted by background and achievement through patterns of residential segregation, and they may also be grouped or tracked within schools. Ignoring this fact is likely to result in penalizing teachers of low-performing students and favoring teachers of high-performing students, just because the teachers of low-performing students cannot move as fast.

Yet another kind of peer effect arises when some students in the classroom directly promote or disrupt the learning of others. Perhaps one or two students were highly disruptive or repeatedly pulled the classroom discussion off topic, wasting precious minutes before the teacher could get the lesson back on track. Simply put, the net result of these peer effects is that VAMs will not simply reward or penalize teachers according to how well or poorly they teach. They will also reward or penalize teachers according to which students they teach and which schools they teach in. (Haertel, 2013, pp. 12–13)

The inconsistency with most VAMs is that the results are biased by student-level variables that are not factored into the model. This affect not only teachers who teach students with special needs and English language learners but also those who teach gifted students. High-achieving
students cannot demonstrate substantial growth because their score cannot pass a certain proficiency level (ceiling), resulting in a capped measure (Wright, Horn, & Sanders, 1997).

The use of value-added measures poses concerns for many researchers when used in schools that are not considered “typical” and have varying levels of student achievement. The achievement gap affecting minorities and lower socioeconomic students within schools continues to be a topic of conversation among education leaders and policymakers. The conversation revolves around accountability with the notion that all students should be able to pass state assessments when all teachers are using the same state standards. The percentage of students who pass the state assessments gives a sense of how high or how many students reached an achievement level but does not measure how much growth of learning took place within a specific timeframe. Moreover, it is difficult to measure students’ learning growth knowing that they all start at different levels with different socioeconomic backgrounds (community characteristics), with different teacher characteristics (e.g., teacher preparation, years of experience, qualifications), and school characteristics (e.g., leadership, resources, interventions, policies). Franco and Seidel (2014) examined the impact of value-added approaches and teacher ratings in schools that were not identified as “typical.” Many urban schools are not typical in the sense that they have achievement gaps among student groups and often have student demographics and teacher characteristics that are different than the typical schools. VAMs are intended to measure student achievement within a given timeframe which is attributed to teachers working with students in particular classrooms and school buildings. Researchers have noted that student academic progress is influenced by student-, teacher-, and school-level variables. Some variables that influence student progress include the student’s prior year achievement level, motivation, and socioeconomic factors (e.g., poverty level, parental
education) that are strongly linked to achievement score results.

Even within schools and grade levels, researchers have identified some valid concerns about using VAMs. Amrein-Beardsley and Collins (2012) pointed out in a study conducted in the Houston Independent School District that almost 46% of teacher evaluation ratings changed from effective to ineffective or vice versa when teachers moved to different grade levels. The study by Darling-Hammond et al. (2012) examined teacher evaluation data from five school districts and found that 20% to 30% of the teachers who were rated less effective in one year were rated the same the following year. Furthermore, 25% to 45% of the teachers rated less effective moved to the highly effective rating the following year. The same was true for those who were rated highly effective in one 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 yield inconsistent patterns of results for individual teachers over time. 2. Teachers’ value-added performance is affected by the students assigned to them in a given year, thereby calling into question the transparency and fairness of using value-added measures of student learning in evaluations. 3. Value-added ratings are unable to disentangle the many other influences that contribute to student progress, thereby providing an incomplete and distorted measure of an individual teacher’s effectiveness. Most importantly, research reveals that gains in student achievement are influenced by much more than any individual teacher. Others 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)
Current Concerns Regarding Value-Added Models

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 model, the Value-Added Research Center (VARC) model, and homegrown models. In four states (including New Jersey), teacher consequences were 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 and heavily influenced by growth or value-added scores. 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 was yet to be determined.

Many states had apprehensions around the inability to use growth and value-added measures with teachers of non-tested grades. All the states currently calculating valued-added scores used state standardized test scores in grades 4–8, which account for 30% of the teacher evaluations. This was of concern for states that used these measures to make consequential decisions. States will find it difficult to make evaluative comparisons within and across schools when most teachers are not evaluated with value-added scores. In terms of reliability, some states expressed concerns with the current research that has indicated a lack of reliability across
the growth models and VAMs. There were mixed concerns related to validity. Some were concerned with the accuracy of the data used when linking students to the teacher of record, and some expressed concerns on whether their state assessments were appropriately designed to measure teacher effectiveness over time. Darling-Hammond (2015) suggested that:

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 the new national assessments:

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. Sanders and Rivers (1996) and Jordan et al. (1997) examined teacher effects on student achievement on statewide assessments in Tennessee and Dallas, Texas. Their studies found that teacher effectiveness was a strong determinant of student
learning. Students who were placed with high-performing teachers three years in a row scored at the 96th percentile on the mathematics state assessment. They performed better than students who were placed with low-performing teachers three years in a row and scored at the 44th percentile on the same assessment. Many of the studies have not specifically identified the characteristics and classroom practices that were linked to effective teachers which would likely improve student learning (Goe, 2007). Teacher effectiveness has been measured in broad terms. Some of the primary teacher characteristic (variables) that have been examined are teacher qualifications, teacher experience, teacher attributes, and teacher practices.

Teacher qualifications include their credentials, certifications, years of experience, subject matter taught, and degrees earned. Goe (2007) examined many research studies that have linked teacher characteristics to student achievement. The findings indicated that teacher qualifications were consistently associated with increased student achievement in mathematics at all grade levels, but more so at the secondary level. Students taught by teachers with stronger mathematics knowledge performed better than students taught by less-knowledgeable mathematics teachers. 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 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. They found that students of teachers who had an undergraduate or graduate degree in mathematics performed better than students whose teachers did not have a mathematics
degree by a small margin of 0.08 standard deviation. 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.

Cavalluzzo’s (2004) research examined the effectiveness of teachers with National Board
Certification (NBC) on student achievement. In this research, nearly 108,000 individual student
records were collected from Miami-Dade County Public Schools to measure the contribution that
teacher characteristics made to student achievement in mathematics grades 9 and 10. The teacher
characteristics tested in the model were years of experience, advanced degree held,
undergraduate school attended, regular state certification in middle school and high school
mathematics, teaching position in mathematics or another primary job assignment, and NBC
status. The study controlled for student characteristics which included age and grade level,
grades repeated, gifted or not, suspension and attendance record, grade point average in core
subjects, average scores in mathematics for effort and for conduct, student achievement above or
below grade level, and enrollment in a limited-English-proficiency program. The findings
indicated a statistically significant contribution to student outcomes for each of the teacher
characteristics except undergraduate school attended. Teaching experience marginally improved
middle school student achievement in mathematics and reading. Cavalluzzo (2004) noted:

Teachers with National Board Certification had an effect size of 0.074 when compared to
otherwise similar teachers. Students who have a teacher with a regular state certification
in high school mathematics have an expected effect size gain of 0.057. In other words,
this credential adds 5.7 percent of a standard deviation to test scores for otherwise
identical students. Teachers in pay step 3 or above have an effect size of 0.05 when
compared to similar teachers who are at pay step 1 or 2. Having a teacher with regular
state certification in middle school mathematics or a graduate degree has smaller effects
on student outcomes. (p. 27)
Goldhaber and Anthony (2005) examined the relationship between NBC and student achievement. The study linked 32,399 teachers to 609,160 students’ reading test scores and linked 32,448 teachers to 611,517 mathematics test scores in North Carolina. The findings were marginally statistically significant. There were student achievement gains for students whose teachers had completed NBC by 0.05 standard deviation in reading and 0.09 standard deviation in mathematics. The findings in both studies were significant but were not of practical importance. Furthermore, teachers with NBC were more than likely teaching affluent students.

Hanushek, Kain, O’Brien, and Rivken (2005) compiled a study to determine the association between teacher certification exam scores, educational attainment, teacher race, years of experience, and student achievement in mathematics on the Texas Assessment of Academic Skills. The data was from a large urban district that included approximately 230,000 student records in grades 4–8 during years between 1989–1990 and 2001–2002. The study found that teacher experience predicted higher student achievement gains in the first few years of teaching and that advanced degrees and certification exam scores were unrelated to student achievement. In addition, they found that a match between student and teacher race improved achievement scores for minority students only.

The racial pairing of teachers and students to determine student achievement has shown mixed results in a few research studies. Dee (2004) conducted a study examining teacher race and student test scores from Tennessee’s Project STAR Public Access Data. The study showed that student assignment to an own-race teacher significantly increased the math and reading achievement of both Black and White students. For Black students, having a Black teacher for one year was correlated with 3- to 5-percentile-point increases in mathematics achievement and 3- to 6-percentile-point increases in reading. Similarly, White students placed with a White
teacher scored 4 to 5 percentile points higher in mathematics and 2 to 6 percentile points higher in reading. The results implied that continued years of students being exposed to own-race teachers had additive effects to student achievement over time. However, the study did not provide evidence on the specific teacher qualities that influenced student achievement. An older study by Ehrenberg, Goldhaber, and Brewer (1995) examined data from the National Education Longitudinal Study of 1988 to determine whether teacher race, gender, and ethnicity contributed to student achievement. They found that the match between teachers’ race, gender, and ethnicity and those of their students had little association with student achievement.

Some 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. Holtzapple (2003) compared student achievement with teachers’ evaluation scores that derived from Danielson’s Framework for Teaching. He studied the evaluations of 246 Cincinnati Public School teachers in grades 3–8 that were linked to student achievement. The study found that teachers who received low ratings on the instructional domain of Danielson’s instruments had students with lower achievement scores. Conversely, teachers with advanced or distinguished ratings generally had students with higher than expected test scores, and teachers rated proficient had students with average gains. Milanowski (2004b) conducted a similar 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. He 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.
Rivkin, Hanushek, and Kain (2005) sought to explain the impact of teachers and schools on student achievement gains on the Texas Assessment of Academic Skills in reading and mathematics. The study examined teacher characteristics that were observable (teacher education and experience) and unobservable components and their relationship to student achievement. The study focused on grades 3–7, and student scores ranged from 143,314 to 455,438. The study found that observable teacher characteristics had marginal but significant effects on student achievement gains. However, most teacher effectiveness was due to unobservable differences in instructional quality, and the study found that teacher effectiveness increased during the first year but leveled off after the third year.

School Characteristics and the Impact on Student Achievement

Okpala, Smith, Jones, and Ellis (2000) examined the impact of school, teacher, and student 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 characteristics identified in the study were average class size, school size, percent of teachers with master’s degrees, percent of teachers with more than 10 years’ teaching experience, percent of students on free or reduced lunch, percent of parents with post-high-school education, and parent volunteer hours. The results from the study indicated that class and school size were significant in explaining achievement gains in reading only and teachers with master’s degrees were significant in mathematics only. Additional findings indicated a significant correlation between teachers with 10 years of teaching experience and student achievement in mathematics and reading. The percentage of parents with post-high-school education was positively correlated in both mathematics and reading achievement. Conversely, there were no significant findings
between student achievement and the percentage of students on free or reduced lunch and parental volunteer hours.

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 high ratings on a school audit instrument developed by the state. 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.

**Summary**

The literature has indicated that classroom teachers impact student achievement and that the variance between classrooms may depend on the quality or effectiveness of the teacher in providing instruction. Studies have supported the concept of holding teachers accountable to student learning, but the debate is still brewing regarding how to do this in a fair and equitable way. Findings have supported the use of VAMs to measure teacher effectiveness by controlling for factors that are outside the teacher’s influence using prior-year assessment data compared against the current year to measure the value of learning added during the year. VAMs have their place in measuring teacher effectiveness, and there is a national push to incorporate them in teacher evaluations. However, VAMs do not have the capability to measure teacher qualities that may contribute to student learning and should be accounted for when evaluating teachers. Qualities such as enthusiasm, verbal ability, flexibility, and creativity are some of the variables
that all effective teachers possess. It can be assumed that effective teachers embed these qualities into their teaching practices which may result in student academic growth. The research literature has identified 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 may alter teaching habits where teachers may teach to the test and ignore other curricular lessons that may be tested in the following school year, hence putting the next-year teachers at risk. The research findings have identified how VAMs may not detect the effectiveness of teachers when teaching low-performing students and high-performing students. The use of proficiency cut scores to determine teacher effectiveness may mask the growth of students who are low-performing and did not meet proficiency levels but demonstrated substantial growth as well as high-performing students who perform at the top of the achievement scale and demonstrate minimal levels of growth. Sanders (2000) suggested that there is a pattern that exists in inner city schools where low-performing students have more opportunities to make reasonable growth. Conversely, high-achieving students in the same schools are being held to the same pace as the lower-achieving students. When this pattern is repeated over years, the high-achieving students lose ground. This may not be fair to the overall process of rating teachers using only value-added measures. The true learning growth of the high-achieving student can go undetected since there is a “ceiling” on how much growth can be measured. This is a disadvantage for teachers who will demonstrate minimal growth when a value-added score is calculated.

The use of VAMs may discourage teachers from working in low-performing schools or with high-need students, which in fact will make these schools and classrooms harder to fill with certified teachers. Darling-Hammond (2015) reported that:
teachers have noted their value-added scores go down when they are assigned to teach in fourth grade where English learners are transitioned into mainstreamed classrooms, and this dip leads to dismissals. One teacher commented, “I’m scared I might lose my job if I teach in a transition grade level, because … my scores are going to drop.” Another explained, “When they say nobody wants to do 4th grade—nobody wants to do 4th grade. Nobody!” (Amrein-Beardsley & Collins, 2012, p. 16 as cited in Darling-Hammond, 2015. p. 134).

The research literature has highlighted a wide range of issues related to the validity and reliability of VAMs. The degree in fluctuation from year to year in teacher ratings is quite concerning. Teachers can be rated effective one year and ineffective the following year. The VAMs are sensitive to statistical analysis and change the outcomes depending on the statistical model being used. There are many questions regarding the validity of state tests and their ability to measure teacher effectiveness. 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 address the extent to which high- versus low-performing teachers differ in their instructional practices, use of questioning, and classroom management practices that result in increased student achievement. They do not consider some of the most vital components that impact student learning such as class sizes, curriculum materials, instructional time, availability of specialists and tutors, home and community supports or challenges, or summer learning loss. Nonetheless, almost all states are employing, piloting, or developing growth models and VAMs to help measure teacher effectiveness. The variability from state to state demonstrates that there is still much to learn and develop to hold teachers accountable for the learning of students. Much
more dialogue and research are needed to determine the validity of VAMs to inform practice and improve teacher effectiveness.

The research has suggested that there are some teacher and school characteristics that impact student achievement. Teacher qualifications in mathematics have been positively associated with increased student achievement. Students taught by teachers with stronger mathematics knowledge and with mathematics certifications and degrees perform better than students taught by less-knowledgeable mathematics teachers. Teacher experience has demonstrated a positive impact on student achievement for the first few years of teaching. In addition, linking teachers and students by race indicated improvement in achievement scores, but the results were mixed in some studies, which did not indicate a significant finding. Measuring a teacher’s performance and student achievement based on school-level constructs has its challenges. More so when comparing teacher effectiveness 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. An effective teacher in a suburban affluent school district may fail to be effective in an at-risk school in an urban setting, and vice versa. Hence, school characteristics and teacher practices make important contributions to student success. Teacher effectiveness is relative within the contexts of the schools when we identify the practices effective teachers use in at-risk schools and in affluent districts that ensure high levels of student learning.
CHAPTER III
RESEARCH METHODOLOGY

Introduction

In 2012, New Jersey enacted the TEACHNJ Act and AchieveNJ to reform teacher tenure laws and to link teacher tenure to evaluation ratings. Teachers would no longer be judged by a single result of student proficiency on a state assessment. Under AchieveNJ, multiple measures were used to evaluate teachers with the approach that students enter classrooms at various 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 was to determine the value added by teacher practices, teacher characteristics, and school characteristics on student growth.

This study will help district leaders evaluate the new state mandates put in place for teacher evaluations and will add to the body of research related to teacher practice, specifically in an urban environment with large numbers of poor and low-performing students. The 2013–2014 school year was the first full year of statewide implementation of the new evaluation mandate, allowing 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 academic achievement (growth) and its relationship to teacher practice, teacher characteristics, and school characteristics; and (d) additional contribution to student growth made by teacher practice over and beyond those associated with teacher and school characteristics.

The 2013–2014 school year was also the first full year that the state of New Jersey provided SGP scores for students in the content areas of mathematics and language arts. An SGP...
describes a student’s growth relative to their academic peers who had the same NJ ASK scores for the past 3 years. Students were measured against their peers to determine academic growth and provided SGP scores categorized as low, typical, or high growth. To determine growth for a group of students within a course or class, student SGP scores were listed in ascending order and the mSGP for the class was assigned to the teacher as a score. For the purpose of this study, the teacher’s mSGP 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 because it provided a structure to collect data and answer the research questions. Quantitative research is an approach of inquiry used to answer questions about relationships among measured variables, with the purpose of explaining, predicting, and controlling phenomena. It is also defined as the collection and analysis of numerical data to describe, explain, or predict a phenomenon of interest (Gay, Mills, & Airasian, 2009). This study sought to explain the relationships between variables that predict student academic growth. The study took a practical approach in these analyses by using mSGP scores that schools received from the state annually, along with 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 could be captured at only one point in time during the 2013–2014 school year. This was the first full year of implementation for the AchieveNJ mandate in which teacher mSGP scores (student growth) in language arts and mathematics were calculated and reported as a multiple measure for teacher evaluation purposes. Furthermore, it was the last year the New Jersey Department of Education administered the standardized statewide NJ ASK assessments in language arts and mathematics. This study aimed to examine 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 teachers with a valid mSGP score who taught in grades 4–7 in the content areas of language arts or mathematics. The study consisted of 30 schools with different grade configurations as indicated in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Grade configurations</th>
<th>P-5</th>
<th>K-4</th>
<th>K-5</th>
<th>K-8</th>
<th>5-8</th>
<th>6-8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of schools</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>21</td>
<td>1</td>
<td>2</td>
<td>30</td>
</tr>
</tbody>
</table>

The New Jersey Department of Education categorized schools within districts as Priority, Focus, or Reward schools based on schoolwide proficiency levels measured by the state assessment, NJ ASK. A Priority school was identified as among the lowest-performing five percent of schools in the state of New Jersey. Focus schools included the overall lowest subgroup performance and the widest achievement gaps between different subgroups of students. A
Reward school achieved high proficiency levels including progress toward closing the achievement gap. There were no Reward schools in the district studied; however, there were schools identified as NonStatus schools which are performing at or above achievement levels. For this study, there were 6 Priority schools, 14 Focus schools, and 10 NonStatus schools. The sample population consisted of 149 language arts ($n = 149$) and 145 mathematics ($n = 145$) teachers in grades 4–7. Each teacher in the study received an mSGP score of their class or course roster. The 294 teacher mSGP scores were reflective of 7,220 students who received a language arts SGP score and 7,163 students who received a math SGP score. The study excluded special education and bilingual teachers since there were few valid mSGP scores, and the study did not control for variables such as: pull out and push in support, the number of years in program, and type of classification.

**Setting**

The study took place in a large urban school district that enrolled over 28,000 students from preschool through grade 12. The district’s population was 62% Hispanic, 28% percent African American, and 10% Caucasian, Middle Eastern or Asian descent. Approximately 50% of all students spoke a primary language other than English, with about 37 languages spoken throughout the schools. The district consisted of 54 schools with approximately 2,500 certified teachers. The district was designated by the state of New Jersey as a district “in need of improvement.” It was one of the 31 former Abbott school districts and one of four school districts that were under state control. The New Jersey Supreme Court found that the education provided to school children in poor communities was inadequate and unconstitutional and instructed that state funding for identified Abbott school districts be equal to the funding spent in the wealthiest districts in the state.
The 2013–2014 school year was the first full year of statewide implementation of the new evaluation policy, AchieveNJ. In compliance with the mandate, the district evaluated approximately 2,000 teachers using a state-approved evaluation instrument named Focal Point Teaching Practice Model. Substantial professional development on the new evaluation system was provided for teachers and administrators. Teachers received numerous workshops on utilizing best practices to meet the needs of diverse learners, and administrators received training on providing meaningful feedback and identifying specific evidence of best instructional practices to share with teachers.

The 2013–2014 school year was the first time teacher evaluation scores were tied to student achievement with an mSGP score based on NJ ASK test results. For this study, mSGP data was available only for teachers in grades 4–7 in language arts and mathematics. Grade 8 teachers were omitted from the study because the state did not calculate mSGP in grade 8 mathematics.

**Instrumentation and Variables**

The NJ ASK state assessment has been administered for several years and has met the reliability and validity criteria as indicated in the New Jersey Assessment of Skills and Knowledge, 2013 Technical Manual (NJ ASK Technical Manual, 2014):

The test reliabilities measured by Cronbach alpha for the 2013 NJ ASK are described in Part 8. The alphas for overall student responses ranged from 0.81 to 0.89 for ELA, 0.90 to 0.93 for mathematics indicating that the tests are highly reliable. (p. 28)

Test validity is reflected in a process where:

Measurement Incorporated (MI) followed statistical and content specifications to make sure that the 2013 NJ ASK assessments are valid. The statistical specification described the psychometric characteristics of the items included in the 2013 assessments. The primary statistical targets used for NJ ASK test assembly were the p-value estimates also
called proportion correct or item difficulty, the point bi-serial correlation which is a measure of how well the items discriminate among test takers and is related to the overall reliability of the test, and proportion correct value which is an indication of test difficulty. Similarly, the minimum target value for a proportion-correct was set at 0.25 and maximum was set at 0.95. In addition, content experts made sure that the items selected for the 2013 NJ ASK tests were free from poor model fit and differential item functioning when they were first field tested. (NJ ASK Technical Manual, 2014, p. 144)

The manual explained:

The tests are constructed under same blueprint and specifications is evidence of content validity. The testing items are developed to align and measure the NJ core curriculum standards so that all students can demonstrate the knowledge and skills necessary for the attainment of proficiency in the academic content areas. All standards and assessments are reviewed by specialists from NJ content as well as bias and sensitivity review committees to identify and eliminate elements that may favor one group (e.g., language, culture, ethnicity) over another. Test items are developed under universal test design principle with NJ special student populations in mind so that no student group is disadvantaged. The test validity is also reflected in the fact that the test is inclusive for all students. The test validity further ensures the comparability and interpretation of scores and proficiency standards across different student groups. All NJ ASK item responses for a given grade/content from the general and special populations are combined for item analysis, calibration, and equating. These analyses include all students regardless of the test version taken, i.e., operational, Spanish, Braille, or Large Print. An entirely different score conversion table is prepared for tests requiring modifications such that a subset of the total number of items constitutes the total score. However, these special test versions are placed on the same scale as the operational tests; thus, proficiency standards can be applied uniformly to all tests. (NJ ASK Technical Manual, 2014, p. 29)

The mathematics portion of the NJ ASK measured a student’s ability to solve several mathematical concepts such as number and numerical operations, geometry and measurement, patterns and algebra, data analysis, probability and discrete mathematics. The mathematics test consisted of multiple-choice and open-ended questions. Open-ended questions required students to apply their knowledge by providing enough information to solve the problem. The open-ended questions were scored on a scale from 0 to 3. The English language arts NJ ASK tests measured students’ reading and writing knowledge. The test included reading passages, multiple-choice questions, constructed-response items, and writing tasks. Writing prompts differed by grade and
included an informative/explanatory prompt, a narrative prompt, and/or a persuasive prompt. Writing prompts were scored on a 0–4 point rubric.

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

**Teacher mSGP Scores**

Student growth was measured by using the teacher’s mSGP score provided by the state’s Department of Education for each teacher in grades 4–7 in language arts and mathematics. To determine the mSGP for an individual teacher, the teacher’s class or course roster was used to create an ascending list of SGP scores for students assigned to the teacher’s class. An SGP describes a student’s growth relative to his or her academic peers who had the same state assessment (NJ ASK) scores for the past 3 years. 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 was reported as an SGP and specified on a scale from 1 to 99 how an individual student’s growth compared to his or her academic peers. A student’s SGP growth was categorized as low (SGP < 35), typical (SGP > 34 and SGP < 66), or high (SGP > 65). Teacher mSGP scores were used as a dependent variable for this study and coded as 0 (low), 1 (typical), or 2 (high); in a logistic analysis, 0 represents low scores while 1 equals typical or high scores.
Teacher Characteristics

Model 1 used an ordinal regression analysis to test the relationship between the predictor variable teacher characteristics and student growth (dependent variable) to determine if teacher characteristics such as subject taught, grade level assignment, and gender and ethnic background were significant predictors of student growth. Each teacher characteristic was coded as follows:

- **subject taught**: 0 = language, 1 = mathematics,
- **grade level taught**: 0 = elementary, 1 = middle,
- **gender**: 0 = female, 1 = male
- **ethnicity**: 0 = Hispanic, 1 = all others
  - 0 = White, 1 = all others
  - 0 = Asian/Pacific Islanders, 1 = all others.

School Characteristics

Model 2 used an ordinal 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 (Priority, Focus, or NonStatus schools) and the student ethnicity within the school influenced student growth. School characteristics were coded based on school performance status:

- 0 = Priority, 1 = Focus and NonStatus
- 0 = Focus, 1 = Priority and NonStatus
- 0 = NonStatus, 1 = Priority and Focus
The ethnic subgroup characteristics addressed the percentage of each ethnic subgroup enrolled in each school:

Black enrollment, \(0 = 20\% \text{ and less than, } 1 = 21\% \text{ and greater than}\)

Hispanic enrollment, \(0 = 59\% \text{ and less than, } 1 = 60\% \text{ and greater than}\)

White enrollment, \(0 = 2\% \text{ and less than, } 1 = 3\% \text{ and greater than}\)

Asian enrollment, \(0 = \text{less than } 1\%, \ 1 = 1\% \text{ and greater}\)

**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 standards and weighted as specified by the Focal Point Teaching Practice Model which was adopted by the district and approved by the state’s Department of Education to observe teacher practice. The framework focused on the following seven performance criteria for teachers: (a) preparation for instruction, (b) use of data to inform instruction, (c) delivery of instruction, (d) interventions to meet diverse needs, (e) classroom environment, (f) leadership, and (g) professionalism. Teacher practice scores were coded and categorized as 1 (ineffective), 2 (partially effective), 3 (effective), or 4 (highly effective). For this study, teacher practice was coded as 0 (ineffective) or 1 (effective). Table 3 lists the variables and measurements that were conducted in the study.
Table 3

**Instrumentation and Variables**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Measurement</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher mSGP (Student Growth)</td>
<td>mSGP: 0 = Low (mSGP &lt; 35) (ordinal regression analysis)</td>
<td>Dependent</td>
</tr>
<tr>
<td></td>
<td>mSGP: 1 = Typical (mSGP &gt; 34 &lt; 66) (ordinal regression analysis)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mSGP: 2 = High (mSGP &gt; 65) (ordinal regression analysis)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mSGP: 0 = Low Growth (used in logistical analysis)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 = Typical and High (used in logistical analysis)</td>
<td></td>
</tr>
<tr>
<td>Teacher Characteristics</td>
<td>Subject taught: 0 = Language 1 = Mathematics</td>
<td>Independent</td>
</tr>
<tr>
<td></td>
<td>Grade level taught: 0 = Elementary 1 = Middle</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender: 0 = Female 1 = Male</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ethnicity: 0 = Black 1 = All others</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = Hispanics 1 = All others</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = White 1 = All others</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = Asian/Pacific Islanders 1 = All others</td>
<td></td>
</tr>
<tr>
<td>School Characteristics</td>
<td>School performance status: 0 = Priority 1 = Focus and NonStatus</td>
<td>Independent</td>
</tr>
<tr>
<td></td>
<td>School performance status: 0 = Focus 1 = Priority and NonStatus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>School performance status: 0 = NonStatus 1 = Priority and Focus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Black enrollment percentage 0 = 20% and less than 1 = 21% and greater than</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hispanic enrollment percentage 0 = 59% and less than 1 = 60% and greater than</td>
<td></td>
</tr>
<tr>
<td></td>
<td>White enrollment percentage 0 = 2% and less than 1 = 3% and greater than</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asian enrollment percentage 0 = less than 1% 1 = 1% and greater than</td>
<td></td>
</tr>
<tr>
<td>Teacher Practice</td>
<td>Rating from observation on a scale from 1 to 4 (1 and 2 = Ineffective; 3 and</td>
<td>Independent</td>
</tr>
<tr>
<td></td>
<td>4 = Effective)</td>
<td></td>
</tr>
</tbody>
</table>

**Procedures**

The district granted authorization to conduct the research study. The process for approval included a *District Research Request Application* that needed to be completed by answering seven 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 were gathered from the district’s reporting system and uploaded into the state’s NJSMART portal. The state calculated and provided the SGP scores for individual students and a median SGP score for individual teachers. A comprehensive
data file was downloaded from the NJSMART portal that included the teachers’ practice scores, students’ SGP scores, teachers’ mSGP scores, teacher assigned schools, and teacher assigned students. The district provided a data file that identified each teacher’s characteristics in terms of gender, subject taught, grade level, and ethnicity. School characteristics were collected from the New Jersey Department of Education website that categorized schools in this study as either Priority or Focus schools. Schools not considered Priority or Focus schools were regarded as NonStatus schools in this study. Student ethnicity enrollment by schools was obtained by the school performance reports on the New Jersey Department of Education website.

Data Analysis

In this study, the ordinal and logistic regression methods were used to test the relationships between the dependent variable (student growth) and independent variables (teacher characteristics, school characteristics, teacher practice). This ordinal regression method allowed the researcher to identify the magnitude of independent variables (subject, grade level, gender, ethnicity, school performance, and student ethnicity) that contributed to student growth. The logistic method was used to be predictive to better explain the relationship between the independent (teacher practice, teacher characteristics, school characteristics) and dependent (student growth) variables.

As shown in Figure 2, the design consisted of three separate models used to answer three research questions. An ordinal regression analysis was used to analyze Model 1 (teacher characteristics on student growth) and Model 2 (school characteristics on student growth). In the full model (Model 3), a logistic regression analysis was used to better interpret the impact of teacher practice and teacher and school characteristics on student growth. The dependent variables mSGP dummy coded variables (low, typical, high) were collapsed into two variables.
(low and typical-high), and the significant predictor variables from Models 1 and 2 were included in the model with teacher practice.

Figure 2. Models testing the impact on student growth. This figure illustrates the three separate models used to answer three research questions to determine influences on student growth.

This analysis seeks to understand the value added by teacher practice and teacher and school characteristics on student growth over the course of the 2013–2014 school year. The model presents findings on student growth from multiple angles, specifically on how teacher practice, subject taught, grade level, teacher gender, teacher ethnicity, student ethnicity, and school-level performance relates to student growth. Variables were coded to distinguish differences in characteristics. Using the danielsoper.com website, a hierarchical multiple regression power analysis was conducted, and it was determined that the minimum required sample size be equal to 95. A $p < .05$ level of significance was used for all analyses in the study to determine if the null hypotheses ($r^2$ change) could be rejected between the models. Preliminary analysis was conducted to examine the assumptions of multicollinearity among the independent variables. An independent variable would be omitted if the Statistical Package for
the Social Sciences (SPSS) program gave a variance inflation factor (VIF) of 2.5 or higher. SPSS was used to determine the degree of variance among the models with multiple regression analyses. A detailed analysis of the data, the findings, and conclusion are presented in Chapters IV and V.
CHAPTER IV

RESEARCH FINDINGS

Introduction

The overarching purpose of this study was to understand the relationship between teacher practice and student growth in a large urban district. For several years, the topic of conversation for many education reformers has been the demand for educational accountability in public schools to improve student achievement on standardized tests. In response to the federal Race to the Top initiative, incentives were provided for states that redesigned their evaluation systems.

The new teacher evaluation system in the state of New Jersey represented a departure from prior approaches to teacher evaluations that were procedural and systemic in nature. The new teacher evaluation system connected teacher practice to student achievement. Thus, the traditional methods of evaluating teacher practice through classroom observations were revised to include the calculations of student achievement through district- and/or school-level assessments and state standardized assessments. Past teacher evaluation models typically used checklists with little observational feedback and rarely included data on student achievement (Wise et al., 1985). 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 study in which ordinal and logistic regression analyses were used to answer questions related to how teacher practice predicts student growth. The analyses consisted of testing three models, one research question for each model, which examined the relationship between student academic growth while controlling for
teacher practice and teacher- and school-level characteristics. This chapter presents the findings from these analyses.

The research was guided by the following questions:

1. Are teacher-level variables such as grade level taught, gender, and ethnic background significant predictors of student growth?

2. To what extent do the following school-level variables influence student growth: school performance status (Priority schools, Focus schools, NonStatus schools) and percent of student subgroup ethnic composition?

3. How is student growth impacted by a teacher’s effectiveness as measured by the practice score received, when one controls for teacher- and school-level characteristics?

This chapter presents the major findings. The outcome variable (student growth) was operationalized using teacher mSGP categorized as low, typical, or high growth (New Jersey Department of Education, 2016).

Participant Demographics

Table 4 provides a breakdown of the characteristics of the 294 teachers with valid mSGP scores included in the study. Of the 294 teachers, 51% taught language arts, and 49% taught mathematics; further, 43% were assigned to elementary grades and 57% to middle school grades. Additionally, 82% were female teachers, while 58% were White, 25% were Black, 13% were Hispanic, and 4% were Asian/Pacific Islander.
Table 4

**Teacher Demographics: Subject Taught, Grade Span, Gender, Ethnicity**

<table>
<thead>
<tr>
<th>Teacher Subject Taught</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = LAL, 1 = Math</td>
<td>Language Arts 149</td>
<td>50.7%</td>
</tr>
<tr>
<td></td>
<td>Mathematics 145</td>
<td>49.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher Grade Span</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Elem; 1 = Middle</td>
<td>Elementary Schools 128</td>
<td>43.5%</td>
</tr>
<tr>
<td></td>
<td>Middle Schools 166</td>
<td>56.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher Gender</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Female; 1 = Male</td>
<td>Female Teachers 240</td>
<td>81.6%</td>
</tr>
<tr>
<td></td>
<td>Male Teachers 54</td>
<td>18.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher Black Ethnicity</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Black; 1 = All Others</td>
<td>Black Teachers 75</td>
<td>25.5%</td>
</tr>
<tr>
<td></td>
<td>Other Teachers 219</td>
<td>74.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher Hispanic Ethnicity</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Hispanic; 1 = All Others</td>
<td>Hispanic Teachers 37</td>
<td>12.6%</td>
</tr>
<tr>
<td></td>
<td>Other Teachers 257</td>
<td>87.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher White Ethnicity</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = White; 1 = All Others</td>
<td>White Teachers 170</td>
<td>57.8%</td>
</tr>
<tr>
<td></td>
<td>Other Teachers 124</td>
<td>42.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher Asian/Pacific Islander Ethnicity</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Asian/PacificIslander; 1 = All Others</td>
<td>Asian/Pacific Teachers 12</td>
<td>4.1%</td>
</tr>
<tr>
<td></td>
<td>Other Teachers 282</td>
<td>95.9%</td>
</tr>
</tbody>
</table>

The school-level variables included in the study are presented in Table 5. There were 30 schools: 6 Priority schools, 14 Focus schools, and 10 NonStatus schools. In terms of teacher placement, 51% of the teachers were from Focus schools, 21% were from Priority schools, and 28% were from NonStatus schools. The teachers were employed in schools with varying student populations. For example, 197 teachers taught in schools with a student subgroup Asian population of 1% or less, 180 teachers taught in schools with approximately 20% or less African-American students, 169 teachers taught in schools with a Hispanic subpopulation of 59% or less, and 168 teachers taught in schools with a White student population of 2% or less.
Table 5

School Demographics: School Performance Designation, Student Ethnicity

<table>
<thead>
<tr>
<th>School Priority Status</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Priority, 1 = Focus and NonStatus</td>
<td>Priority Schools</td>
<td>62</td>
</tr>
<tr>
<td>0 = Focus; 1 = Priority and NonStatus</td>
<td>Focus and NonStatus Schools</td>
<td>232</td>
</tr>
<tr>
<td>School Focus Status</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>0 = Focus, 1 = Priority and NonStatus</td>
<td>Priority Schools</td>
<td>149</td>
</tr>
<tr>
<td>School NonStatus Status</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>0 = NonStatus, 1 = Priority and Focus</td>
<td>NonStatus Schools</td>
<td>83</td>
</tr>
<tr>
<td>School Black Enrollment by Racial Subgroup</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>0 = less than 21%; 1 = greater than 20%</td>
<td>20% and less</td>
<td>180</td>
</tr>
<tr>
<td>21% and greater</td>
<td>114</td>
<td>38.8%</td>
</tr>
<tr>
<td>School Hispanic Enrollment by Racial Subgroup</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>0 = less than 60%; 1 = greater than 59%</td>
<td>59% and less</td>
<td>169</td>
</tr>
<tr>
<td>60% and greater</td>
<td>125</td>
<td>42.5%</td>
</tr>
<tr>
<td>School White Enrollment by Racial Subgroup</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>0 = less than 3%; 1 = greater than 2%</td>
<td>2% and less</td>
<td>168</td>
</tr>
<tr>
<td>3% and greater</td>
<td>126</td>
<td>42.9%</td>
</tr>
<tr>
<td>School Asian Enrollment by Racial Subgroup</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>0 = less than 1%; 1 = greater than 0%</td>
<td>less than 1%</td>
<td>197</td>
</tr>
<tr>
<td>1% and greater</td>
<td>97</td>
<td>33.0%</td>
</tr>
</tbody>
</table>

Regarding student performance, the findings reported in Table 6 indicate that 17% of the teachers had student growth designated as low, 72% demonstrated typical growth, and 11% high growth. The table shows the teacher practice scores received from observations. Most teachers were rated effective. Specifically, while 15% were rated ineffective, approximately 85% received an effective teacher practice rating.

Table 6

Teacher Evaluation Rating: Median Student Growth Percentile, Teacher Practice

<table>
<thead>
<tr>
<th>Teacher mSGP Growth Band (0 = low, 1 = typical, 2 = high)</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Growth</td>
<td>50</td>
<td>17.0%</td>
</tr>
<tr>
<td>Typical Growth</td>
<td>211</td>
<td>71.8%</td>
</tr>
<tr>
<td>High Growth</td>
<td>33</td>
<td>11.2%</td>
</tr>
<tr>
<td>Teacher Practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ineffective</td>
<td>44</td>
<td>15.0%</td>
</tr>
<tr>
<td>Effective</td>
<td>250</td>
<td>85.0%</td>
</tr>
</tbody>
</table>
Findings Model 1

Research Question 1 asked: Are teacher-level variables such as grade level taught, gender, and ethnic background significant predictors of student growth? Model 1 tested the relationship between teacher characteristics and student growth to determine if teacher-level variables such as subject taught, grade-level assignment, gender, and ethnic background significantly explained the odds of students demonstrating growth. These relationships are depicted in Figure 3 below.

Figure 3. Impact of teacher characteristics on student growth. This figure illustrates the teacher characteristic variables that were tested to determine influence on student growth.

A negative log-log function was used in estimating the model in which the ordered category of student growth was the dependent variable and the teacher characteristics variables (subject taught, grade-level assignment, gender, and ethnicity) were the independent variables.

To facilitate understanding of the findings, the dummy codes for each variable are listed below:

**Teacher Characteristics:**

subject taught: 0 = language 1 = mathematics;
grade taught: 0 = elementary 1 = middle;
gender: 0 = female 1 = male;
ethnicity: 0 = Black 1 = all others; 0 = Hispanic 1 = all others;
0 = White 1 = all others; 0 = Asian/Pacific Islander 1 = all other.
Tables 7 and 8 present the model fitting statistics which indicated that the inclusion of the teacher characteristics variables significantly improved upon the intercept only model (Chi Square = 40.79, df = 6, $p < .000$). Moreover, the test of parallel lines results confirmed the assumption that the slope of the coefficients was the same across response categories (Chi Square = 10.803, df = 6, $p < .095$). Based on the Nagelkerke Pseudo R Square reported in Table 9, approximately 16.4% of the variance in student growth was explained by the teacher characteristic variables included in the model.

The parameter estimates in Table 10 indicate that four variables were significant. These variables were: subject taught (language arts), grade level (elementary), gender (male), and ethnicity (Black). With respect to subject taught, the odds ratio of 0.701 was significant at the .023 level (Wald = 5.140, $p < .023$, [CI = -0.661 - -0.048]). Students taught by teachers of language arts were more likely to have typical or high growth than students taught by teachers of mathematics. Regarding grade level, the odds ratio of 0.441 was significant (Wald = 26.605, $p < .000$, [CI = -1.127 - -0.507]). Teachers assigned to the elementary grades were more likely to have students whose growth was either typical or high compared to teachers working in the middle grades. The odds ratio for gender was 1.565 and significant (Wald = 4.769, $p < .029$, [CI = 0.046 - 0.849]). Students taught by male teachers were 1.5 times more likely to have higher growth than students taught by female teachers. Finally, the odds ratio (.405) for the dummy coded variable of teacher race, where African-American was the reference group, was found to be significant at the .048 level (Wald = 3.911, $p < .048$, [CI= - 1.802 - -0.008]). Students taught by teachers who identified as African-American were more likely to have typical or high growth compared to those taught by teachers of other ethnic races.
Table 7

Model 1 Model Fitting Information

<table>
<thead>
<tr>
<th>Model</th>
<th>-2 Log Likelihood</th>
<th>Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept Only</td>
<td>160.146</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>119.347</td>
<td>40.799</td>
<td>6</td>
<td>.000</td>
</tr>
</tbody>
</table>

Link function: Negative Log-log.

Note. df = degrees of freedom

Table 8

Model 1 Test of Parallel Lines

<table>
<thead>
<tr>
<th>Model</th>
<th>-2 Log Likelihood</th>
<th>Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis</td>
<td>119.347</td>
<td></td>
<td></td>
<td>.095</td>
</tr>
<tr>
<td>General</td>
<td>108.544</td>
<td>10.803</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

Link function: Negative Log-log.

Note. df = degrees of freedom

Table 9

Model 1 Pseudo R-Square

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox and Snell</td>
<td>.130</td>
</tr>
<tr>
<td>Nagelkerke</td>
<td>.164</td>
</tr>
<tr>
<td>McFadden</td>
<td>.088</td>
</tr>
</tbody>
</table>

Link function: Negative Log-log.
Table 10

Parameter Estimates for Impact of Teacher Characteristics on Student Growth

<table>
<thead>
<tr>
<th>Location</th>
<th>Subject Matter Language Arts</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Odds Ratio</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Grade Span Elementary</td>
<td>-0.817</td>
<td>0.158</td>
<td>0.4417</td>
<td>26.605</td>
<td>1</td>
<td>0.000</td>
<td>-1.127</td>
<td>-0.507</td>
</tr>
<tr>
<td>Gender Male</td>
<td>0.448</td>
<td>0.205</td>
<td>1.565</td>
<td>4.769</td>
<td>1</td>
<td>0.029</td>
<td>0.046</td>
<td>0.849</td>
</tr>
<tr>
<td>Race African-American</td>
<td>-0.905</td>
<td>0.458</td>
<td>0.4045</td>
<td>3.911</td>
<td>1</td>
<td>0.048</td>
<td>-1.802</td>
<td>-0.008</td>
</tr>
<tr>
<td>Race Hispanic</td>
<td>-0.724</td>
<td>0.480</td>
<td>0.4848</td>
<td>2.277</td>
<td>1</td>
<td>0.131</td>
<td>-1.664</td>
<td>0.216</td>
</tr>
<tr>
<td>Race White</td>
<td>-0.619</td>
<td>0.443</td>
<td>0.5384</td>
<td>1.954</td>
<td>1</td>
<td>0.162</td>
<td>-1.487</td>
<td>0.249</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom

**Findings: Model 2**

Research Question 2 asked: To what extent do the following school-level variables influence student growth: school performance status (Priority schools, Focus schools, NonStatus schools) and percent of student subgroup ethnic composition? Model 2 tested the relationship between school characteristics and student growth to determine to what degree school-level variables such as school performance status (Priority, Focus, NonStatus schools) and student racial composition influenced student growth. These relationships are depicted in Figure 4 below.

Figure 4. Impact of school characteristics on student growth. This figure illustrates the school characteristic variables that were tested to determine influence on student growth.
Like the previous model, a negative log-log function was used in estimating the model in which the ordered category of student growth was the dependent variable and school characteristics the independent variables. The dummy coded variables that were used for school characteristics were:

1. **School performance status:**
   - $0 = \text{Priority}, 1 = \text{Focus and NonStatus}$;
   - $0 = \text{Focus}, 1 = \text{Priority and NonStatus}$;
   - $0 = \text{NonStatus}, 1 = \text{Priority and Focus}$

2. **Ethnic Subgroup:**
   - Black enrollment percentage $0 = 20\%$ and less than, $1 = 21\%$ and greater than
   - Hispanic enrollment percentage $0 = 59\%$ and less than, $1 = 60\%$ and greater than
   - White enrollment percentage $0 = 2\%$ and less than, $1 = 3\%$ and greater than
   - Asian enrollment percentage $0 = \text{less than 1}\%, 1 = \text{1}\%$ and greater than

Tables 11 and 12 indicate that the inclusion of the teacher characteristics variables significantly improved upon the intercept only model (Chi Square = 21.65, df = 7, $p < .003$). Moreover, the test of parallel lines results confirmed the assumption that the slope of the coefficients was the same across response categories (Chi Square = 2.683, df = 7, $p < .913$). Based on the Nagelkerke Pseudo R Square reported in Table 13, approximately 9.0% of the variance in student growth was explained by the school characteristics variables included in the study. Indeed, these variables explained less than the teacher characteristics variables did. The parameter estimates in Table 14 indicate that four school characteristic variables (Priority schools, Focus schools, and Black and Hispanic student population percentages) were significant. The odds ratio for school
designation priority was 1.931 and significant at the .000 level (Wald = 3552.51, \( p < .000 \), \([CI = -15.968 - -14.951]\)). The odds of attaining typical or high growth for students in Focus and NonStatus schools were almost twice as great as those for students taught in Priority schools. With respect to Focus schools, the odds ratio of 1.241 was significant (Wald = 54.5359, \( p < .000 \), \([CI= -16.324- -15.480]\)), implying that students taught in Priority and NonStatus schools were more likely to have typical or high growth than students taught in Focus schools. Conjoining these two results, teachers in NonStatus school settings, as to be expected, were likely to see their students’ growth meeting or exceeding expectations.

Regarding school ethnic compositions, when examining the impact of settings in which the reference category was schools where the proportion of students of African-American background was .20 or less, a significant odds ratio of 1.754 (Wald = 4.475, \( p < .034 \), \([CI=0.046 - 0.849]\)) was obtained. Schools with a Black student population greater than 20\% were likely to demonstrate higher growth than schools with a Black student population less than 21\%. Also, the Hispanic odds ratio was 1.513 and significant at the .050 level (Wald = 3.832, \( p < .050 \), \([CI = -0.001 - 0.907]\)). Students of teachers in schools with a Hispanic student population of 60\% and higher were 1.5 times more likely to have typical or high growth than schools with a Hispanic student population less than 60\%.

### Table 11

<table>
<thead>
<tr>
<th>Model</th>
<th>-2 Log Likelihood</th>
<th>Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept Only</td>
<td>116.336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>94.721</td>
<td>21.615</td>
<td>7</td>
<td>.003</td>
</tr>
</tbody>
</table>

Link function: Negative Log-log.

Note. df = degrees of freedom
Table 12

*Model 2 Test of Parallel Lines*\(^a\)

<table>
<thead>
<tr>
<th>Model</th>
<th>-2 Log Likelihood</th>
<th>Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis</td>
<td>94.721</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>92.038</td>
<td>2.683</td>
<td>7</td>
<td>.913</td>
</tr>
</tbody>
</table>

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.
Note. df = degrees of freedom

Table 13

*Model 2 Pseudo R-Square*

<table>
<thead>
<tr>
<th>Link function: Negative Log-log.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox and Snell</td>
</tr>
<tr>
<td>Nagelkerke</td>
</tr>
<tr>
<td>McFadden</td>
</tr>
</tbody>
</table>
Table 14

Model 2 Parameter Estimates

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>Odds Ratio</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHOOL Priority Status 0 = Priority; 1 = Focus and NonStatus</td>
<td>-15.460</td>
<td>0.259</td>
<td>1.931</td>
<td>3552.525</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>SCHOOL Focus Status 0 = Focus; 1 = Priority and NonStatus</td>
<td>-15.902</td>
<td>0.215</td>
<td>1.241</td>
<td>5453.592</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>SCHOOL NonStatus 0 = NonStatus; 1Priority and Focus</td>
<td>-15.666</td>
<td>0.000</td>
<td>1.572</td>
<td>1</td>
<td>0.034</td>
<td>95% Confidence Interval</td>
</tr>
<tr>
<td>School Black Enrollment by Racial Subgroup 0 = less than 21%; 1 = greater than 20%</td>
<td>0.562</td>
<td>0.266</td>
<td>1.754</td>
<td>4.475</td>
<td>1</td>
<td>0.050</td>
</tr>
<tr>
<td>School Hispanic Enrollment by Racial Subgroup 0 = less than 60%; 1 = greater than 59%</td>
<td>0.453</td>
<td>0.232</td>
<td>1.513</td>
<td>3.832</td>
<td>1</td>
<td>0.122</td>
</tr>
<tr>
<td>School White Enrollment by Racial Subgroup 0 = less than 3%; 1= greater than 2%</td>
<td>-0.261</td>
<td>0.169</td>
<td>0.770</td>
<td>2.392</td>
<td>1</td>
<td>0.706</td>
</tr>
<tr>
<td>School Asian Enrollment by Racial Subgroup 0 = less than 1%; 1= greater than 0%</td>
<td>0.106</td>
<td>0.281</td>
<td>1.111</td>
<td>0.142</td>
<td>1</td>
<td>0.445</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom

Findings: Model 3

Research Question 3 asked: How is student growth impacted by a teacher’s effectiveness as measured by the practice score received, when one controls for teacher- and school-level characteristics? Model 3 tested whether student growth (mSGP) was impacted by teacher practice, teacher characteristics, and school-level characteristics. These relationships are depicted below in Figure 5.
Figure 5. Impact of teacher characteristics, school characteristics and teacher practice on student growth. This figure illustrates the full model tested to determine the impact of teacher characteristics, school characteristics and teacher practice on student growth.

This model introduced the predictor variable (teacher practice) and tested for its explanatory power when controlling for all the significant variables from the previous models. Adding the teacher practice score (ineffective, effective) to the ordinal regression model impacted collinearity among the various other variables and failed to make interpretation clearer; in fact, it made interpretation more difficult. To test this model, a logistic regression analysis was used to better interpret the impact of teacher practice and teacher and school characteristics on student growth. The dependent mSGP dummy coded variables (low, typical, high) were collapsed into two variables (low and typical-high), and the significant predictor variables from Models 1 and 2 were included in the model with teacher practice.
The dummy code variables that were used for the independent variables were:

1. *Teacher practice*: 0 = ineffective, 1 = effective

2. *Teacher characteristics*: subject taught: 0 = language arts, 1 = mathematics; grade level taught: 0 = elementary, 1 = middle; gender: 0 = female, 1 = male; ethnicity: 0 = Black, 1 = all others; 0 = Hispanics, 1 = all others

3. *School characteristics*: school performance status: 0 = Priority and NonStatus; 0 = Focus, 1 = Priority and NonStatus; 0 = NonStatus, 1 = Priority and Focus; ethnic subgroup: Black enrollment percentage 0 = 20% and less, 1 = 21% and greater; Hispanic enrollment percentage 0 = 59% and less, 1 = 60% and greater

Based on the Nagelkerke Pseudo R Square shown in Table 15, approximately 34% of the variance in student growth was explained by the teacher practice score and teacher and school characteristics variables included in the study. Thus, this full model accounts for the greatest proportion of variance (approximately one-third) in student growth.

The results presented in Table 16 indicate that four variables were significant: the covariates of grade level, Black ethnicity, and Hispanic ethnicity and the main variable of interest, the teacher practice score. Students taught by middle school teachers were 9 times more likely to have typical or high growth than those taught by teachers in the elementary grades (Wald = 26.924, \( p < .000 \), \( \text{Exp}[B] = 9.388 \) [CI = 4.029 – 21.875]). Black student enrollment was significant (Wald = 7.524, \( p < .006 \), \( \text{Exp}[B] = .273 \) [CI = .108 – .690]). Schools with a Black student population of 20% or less were likely to have higher growth than schools with a Black student population greater than 21%. Also, schools with a Hispanic student population less than
60% were more likely to have typical or high growth than schools with a Hispanic student population that was 60% or greater (Wald = 5.191, \( p < .023 \), \( \text{Exp} (B) = .326 \) [CI = .125 - .855]). When student growth was conditioned upon teacher practice and after controlling for teacher- and school-level characteristics, a significant odds ratio was found. Specifically, the odds ratio for this variable (5.113) was significant at the .000 level (Wald = 13.025, \( p < .000 \), \( \text{Exp} [B] = 5.113 \) [CI = 2.108 - 12.405]). Students taught by teachers that were rated effective were 5 times more likely to have typical or high growth than students taught by teachers rated ineffective.

Table 15

*Model 3 Model Summary*

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log likelihood</th>
<th>Cox &amp; Snell R Square</th>
<th>Nagelkerke R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>202.241(^a)</td>
<td>.201</td>
<td>.336</td>
</tr>
</tbody>
</table>

\(^a\) Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.
### Table 16

**Model 3 Variables in the Equation**

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable Description</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
<th>95% CI for EXP(B)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Teacher Grade Span</td>
<td>2.239</td>
<td>.432</td>
<td>26.924</td>
<td>1</td>
<td>.000</td>
<td>9.388</td>
<td>4.029</td>
<td>21.875</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = elem., 1 = middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher Gender</td>
<td>-.647</td>
<td>.488</td>
<td>1.763</td>
<td>1</td>
<td>.184</td>
<td>.523</td>
<td>.201</td>
<td>1.361</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = female, 1 = male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher Black Ethnicity</td>
<td>.331</td>
<td>.399</td>
<td>.688</td>
<td>1</td>
<td>.407</td>
<td>1.392</td>
<td>.637</td>
<td>3.041</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = Black; 1 = all others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher Subject Taught</td>
<td>.505</td>
<td>.371</td>
<td>1.853</td>
<td>1</td>
<td>.173</td>
<td>1.656</td>
<td>.801</td>
<td>3.426</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = LAL, 1 = Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>School Priority Status</td>
<td>.068</td>
<td>.620</td>
<td>.012</td>
<td>1</td>
<td>.913</td>
<td>1.070</td>
<td>.318</td>
<td>3.607</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = Priority, 1 = Focus and NonStatus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>School Focus Status</td>
<td>.396</td>
<td>.518</td>
<td>.586</td>
<td>1</td>
<td>.444</td>
<td>1.486</td>
<td>.539</td>
<td>4.101</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = Focus, 1 = Priority and NonStatus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>School Black Enrollment by Racial Subgroup</td>
<td>-1.299</td>
<td>.474</td>
<td>7.524</td>
<td>1</td>
<td>.006</td>
<td>.273</td>
<td>.108</td>
<td>.855</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = less 21, 1 = greater 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>School Hispanic Enrollment by Racial Subgroup</td>
<td>-1.120</td>
<td>.491</td>
<td>5.191</td>
<td>1</td>
<td>.023</td>
<td>.326</td>
<td>.125</td>
<td>.855</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = less 60, 1 = greater 59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher Practice</td>
<td>1.632</td>
<td>.452</td>
<td>13.025</td>
<td>1</td>
<td>.000</td>
<td>5.113</td>
<td>2.108</td>
<td>12.405</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 = Ineffective, 1 = Effective</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>-.079</td>
<td>.931</td>
<td>.007</td>
<td>1</td>
<td>.932</td>
<td>.924</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summary of Quantitative Findings**

Presented in this chapter were the findings for three models that examined the relationship between student academic growth while controlling for teacher practice and teacher- and school-level characteristics. The findings suggest that some teacher-level variables are significant predictors of student growth. Approximately 16% of the variance in student growth,
as measured by teacher mSGP, was explained by ethnicity, gender, subject taught, and grade level.

The New Jersey Department of Education categorized districts in need of improvement and identified schools within the district as Priority, Focus, or Reward schools based on student performance. The results indicated that 9.0% of the variance in student growth was explained by school-level variables. There were four significant findings for typical or high growth for students that were taught by teachers in Focus and NonStatus (p < .000) and Priority and NonStatus schools (p < .000) when the student enrollment population was greater than the median percentage for Black (20% or less, p < .034) and Hispanic (60% or less, p < .050) students. There were no significant findings for school-level characteristics with White and Asian student populations within schools.

A teacher practice score was derived from an average of observation ratings by standards and weighted as specified by the Focal Point Teaching Practice Model which was adopted by the district and approved by the state Department of Education to observe teacher practice. The framework focused on the following seven performance criteria for teachers: (a) preparation for instruction, (b) use of data to inform instruction, (c) delivery of instruction, (d) interventions to meet diverse needs, (e) classroom environment, (f) leadership, and (g) professionalism. Teachers received a practice score rating of 1 (ineffective), 2 (partially effective), 3 (effective), or 4 (highly effective). To conduct a logistic regression analysis in this study, the four categories of teacher practice effectiveness were collapsed into two ratings of either 0 (ineffective) or 1 (effective). Of the total 294 teachers in the study, 15% were rated ineffective, and 85% were rated effective. The model also controlled for the teacher- and school-level characteristics
mentioned in this study to understand if all three variables were significant predictors of student growth.

The findings indicated that 34% of the variance in student growth was explained by teacher practice and school- and teacher-level characteristics. The parameter estimates in Table 17 indicate that four variables were significant. These were grade level (middle, \( p < .000 \)), Black (\( p < .006 \)) and Hispanic (\( p < .023 \)) student enrollment populations, and teachers rated effective (\( p < .000 \)). It is noteworthy that, once teacher practice was added to the logistic regression analyses, teacher gender (\( p < .183 \)), teacher ethnicity (Black, \( p < .407 \)), subject taught (\( p < .173 \)), Priority (\( p < .913 \)), and Focus schools (\( p < .444 \)) were no longer considered significant. Students demonstrated typical or high growth with effective teachers that taught middle grades in schools with Black and Hispanic student enrollment greater than 20% and 60%, respectively.

Table 17 identifies the significance level of all independent variables within the models.

Table 17

**Variables Level of Significance**

<table>
<thead>
<tr>
<th></th>
<th>Ordinal Regression Model 1</th>
<th>Ordinal Regression Model 2</th>
<th>Logistic Regression Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-Level Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject taught: language arts</td>
<td>0.023</td>
<td></td>
<td>0.173</td>
</tr>
<tr>
<td>Grade span taught: elementary</td>
<td>0.000</td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>Teacher gender: male</td>
<td>0.029</td>
<td></td>
<td>0.184</td>
</tr>
<tr>
<td>Teacher ethnicity: Black</td>
<td>0.048</td>
<td></td>
<td>0.407</td>
</tr>
<tr>
<td><strong>School-Level Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus and NonStatus</td>
<td></td>
<td>0.000</td>
<td>0.916</td>
</tr>
<tr>
<td>Priority and NonStatus</td>
<td></td>
<td>0.000</td>
<td>0.444</td>
</tr>
<tr>
<td>Percent of student ethnic enrollment: Black (greater than 20%)</td>
<td>0.034</td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>Percent of student ethnic enrollment: Hispanic (greater than 60%)</td>
<td>0.050</td>
<td></td>
<td>0.023</td>
</tr>
<tr>
<td><strong>Teacher Practice</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effective</td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>
Interestingly, variables that were significant in Models 1 and 2 were no longer significant at the .05 level: subject taught ($p < .173$), teacher gender ($p < .184$), teacher ethnicity ($p < .407$), Priority schools ($p < .946$), and Focus schools ($p < .444$, respectively). The school-level variable elementary grades did not deviate and remained at the .000 significance level.
CHAPTER V
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Chapter V presents a brief introduction of the unique concept of teacher effectiveness within the research literature, student growth within the study, a summary of the research findings, recommendations for further research, and a conclusion.

The policy logic of linking teacher evaluation to student achievement is grounded in research that has confirmed a direct relationship between teacher effectiveness and student academic success (Gates Foundation, 2013; Jordan et al., 1997; Sanders & Rivers, 1996; Stronge et al., 2011). VAMs attempt to predict the “value” a teacher adds to his or her students’ learning growth and may account for up to 50% of the teacher evaluation in many states. Many school districts use this information to determine a teacher’s retention, promotion, compensation, and tenure. However, research studies have indicated mixed results when comparing teacher evaluations and student performance while using VAMs. One study suggested that teachers’ effectiveness improved during the school year and subsequent school years when VAMs were incorporated into the evaluation process (Taylor & Tyler, 2012). Additional studies found little significance when examining the relationship between student achievement and teacher evaluation ratings. The results suggested large fluctuations in correlations between years and across subjects and grade levels. VAMs have demonstrated a lack of reliability in measuring teacher effectiveness to close the achievement gaps between poor and more advantaged students as well as low- and high-achieving students (Borman & Kimball, 2005; Kimball et al. 2004; White, 2004). The inconsistency with most VAMs is that the results are influenced by student, teacher, and school variables that are not factored into the model. It can be difficult to measure students’ learning growth when students enter classrooms with different socioeconomic
backgrounds and different teacher and school characteristics. Franco and Seidel (2014) examined the impact of value-added approaches and teacher ratings in schools that were not identified as “typical” and indicated the disadvantages of using VAMs. Many urban schools are not typical in the sense that they have achievement gaps among student groups and often have student demographics and teacher characteristics that are different than typical schools. Haertel (2013) highlighted the disadvantages for teachers that teach low-performing students in low-performing schools. Ignoring this fact is likely to result in penalizing teachers of low-performing students and favoring teachers of high-performing students, just because the teachers of low-performing students cannot grow as fast. It can be assumed that other factors may have influenced student achievement that were not measured by the growth model used in the district studied in the research. Darling-Hammond et al. (2012) discussed how gains in student achievement may be influenced by class sizes, curriculum materials, instructional time, availability of specialists and tutors, resources for learning (books, computers, science labs), home and community supports, individual student needs and abilities, health, attendance, and prior teachers and schooling.

The premise of AchieveNJ and the other value-added evaluation models discussed in Chapter II is that teacher effectiveness is a measurement of student academic growth. This study is unique as student growth (SGP) is measured using scores of students with like scores across the state of New Jersey. Students are compared to their academic peers to determine growth regardless of their level of proficiency. Sanders (2000) pointed out that a fair accountability system should measure teacher effectiveness by the rate at which students’ progress regardless of their socioeconomic status and the understanding that all students are at different academic levels and will learn at different paces. It is assumed that teacher effectiveness on student growth can
be measured year to year when students are compared to their academic peers with like assessment results.

The school district examined in this study was considered one of the largest and lowest-performing district in the state of New Jersey and was designated as a district in need of improvement. The New Jersey Department of Education classified this district as being in District Factor Group “A”, the lowest of eight groupings that allowed comparison by common socioeconomic characteristics. Many of the students came from impoverished home environments and attended schools where facilitates were inadequate. Academically, the students consistently lagged behind their more privileged peers on state assessments. It is important to recognize that several other factors may influence student growth not measured in teacher effectiveness. Darling-Hammond et al. (2012) cited such other factors as attendance rate, mobility rate, class size, curriculum material, instructional time, prior teacher schooling, and home and community supports. However, these influences were not factored into this study since this study examined student growth and teacher practice (effectiveness) based on observation practice scores.

This study is unique in the sense that teacher effectiveness and student academic growth were examined by measuring students against similar students with like scores across the state rather than comparing student achievement against districts with a similar socioeconomic status as has been done in the past. Teachers and students in poor community school districts can now be compared to the rate of growth of teachers and students in affluent communities through peer grouping.

The purpose of this study was to examine the relationship between teacher practice and student growth, in an urban school district in the state of New Jersey. The analyses included
testing three models, one research question for each model, which examined the relationship between student academic growth while controlling for teacher practice and teacher- and school-level characteristics. This research was a cross-sectional study in which ordinal regression and logistic regression analyses were used to examine this relationship. The total number of participants in the study was 294 language arts and mathematics teachers in grades 4–7 across 30 schools with several teacher and school-level characteristics.

The overall premise of the state of New Jersey new teacher evaluation system and law, AchieveNJ, was that teachers should not be “evaluated on a single factor or test scores alone, but on multiple measures of both effective practice and student learning” (New Jersey Department of Education, 2015, p. 1). Rather than measuring proficiency by a grade level passing score, a growth methodology was used to calculate student learning, recognizing that students enter each grade level at different starting points and with unique challenges. Student growth was measured by comparing the change in their achievement on state assessment from one year to the next, to others students with the same historical state assessment results (i.e., their academic peers). This approach is distinctive in that it illustrated how similar students with like scores across the state “typically” grow in academic performance, regardless of the school district they attend. At the same time, 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). As with students, teachers were assigned a growth percentile score based on the growth of the students in the class. To determine the growth score for a teacher, a class roster was used to create an ascending list of students’ SGPs scores (New Jersey Department of Education, 2015).
Teachers received the median number as their mSGP score, which can used for evaluative purposes and to compare student growth across grades within the district and other districts from year to year. The teacher’s mSGP score is comparative to the three categories identified for student growth (low, typical, or high).

The research literature has noted a wide range of issues related to the validity and reliability of VAMs that do not control for teacher- and school-level variables and run the risk of masking the true effects of teachers on student achievement (Darling-Hammond, 2015). The three models tested in this study controlled for teacher- and school-level characteristics. The findings from Model 1 suggested that some teacher-level variables were significant predictors of student growth. Approximately 16% of the variance in student growth was explained by the teacher’s ethnicity, gender, subject taught, and grade level. Model 2 controlled for school-level characteristics and indicated that 9.0% of the variance in student growth was explained by school variables where students were taught by teachers in Focus and NonStatus and Priority and NonStatus schools when the student ethnic composition was greater than the median percentage for Black and Hispanic students. Model 3 was the full model tested and used to better interpret the impact of teacher practice and teacher and school characteristics on student growth. The significant variables controlled for in Models 1 and 2 were included in the test to understand the value added by teacher characteristics, school characteristics, and teacher practice on student growth. The findings indicated that 34% of the variance in student growth was explained by teacher practice and school- and teacher-level characteristics. Students demonstrated typical or high growth with effective teachers that were teaching middle grades in schools with Black and Hispanic student enrollment greater than 20% and 60%, respectively.
The findings indicate that teacher practice (effectiveness) makes a difference when one controls for teacher- and school-level variables. In this study, 85% of the teachers were rated effective on their observation practice score, and 93% of the teachers’ mSGP scores were at the typical or high growth level. According to the findings, effective teachers are 5 times more likely to have students with typical or high growth. It is important to note that many researchers have expressed the belief that there are limitations to using VAMs, specifically the inability to control for student, teacher, and school variables that impact teacher effectiveness. Teachers’ effectiveness is relative when SGP is a normative variable, measured across thousands of students across the state of New Jersey with similar scores (Betebenner, 2011). This study adds to the body of research and tests the assumptions, derived from the existing literature, that student academic growth is a result of teacher practice (the central variable of interest in the study), teacher characteristics, and school characteristics. It is important to note that teacher practice is influenced by both teacher and school characteristics. Indeed, one could argue that a teacher’s practice mediates the influence of both sets of variables.

Prior to the 2013–2014 school year, the school district implemented the following initiatives to improve student achievement:

1. **Regional Achievement Center:** Priority and Focus schools are monitored by the state’s Regional Achievement Centers (RACs), which are charged with improving the overall performance of Priority and Focus schools. The state’s Department of Education shifted resources to directly support these schools and partnered with the district to set clear goals for student growth, use data to drive decision-making, and implement turnaround principles. The district in this study and the RAC worked collaboratively to achieve the mutual goal of significantly improving student achievement in Priority and Focus
schools. This included creating an aligned instructional system, building pedagogical capacity among teachers and principals, creating a strong district-level support system for schools, and involving parents and community partners in all aspects of the school’s improvement.

2. *End to Social Promotion:* The district established an end to social promotion policy by enforcing retention for students who did not meet grade-level expectations and provided a required summer intervention program in mathematics and language arts for those students.

3. *New Evaluation System:* The district implemented the new teacher and administrator evaluation systems that connected teacher and administrator performance to student achievement results. Intense training was provided to teachers and administrators on the Focal Point Teaching Practice Model. The goal was to transform the evaluation process to make it more rigorous and accurate, which would differentiate teacher effectiveness by enabling the school leaders to use evaluation information to make better decisions related to tenure, assignments, and non-renewals.

4. *Professional Development:* The district collaborated with the University of Pittsburgh’s Institute for Learning to enhance teacher and principal instructional capacity in language arts and mathematics as well as for English language learners through a series of job-embedded professional development activities.

**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 while controlling for teacher- and
school-level characteristics with the same three research questions presented in this study.

2. A follow-up study should analyze the same teachers in years 2 and 3 to determine if there are any fluctuations in their students’ academic growth scores.

3. A comparative study should examine the relationship between student growth in NonStatus schools compared to student growth in Priority and Focus schools. The study could seek to determine whether the academic interventions implemented in low-performing schools impact student growth as measured by SGP.

4. A similar study should be replicated that includes other state-controlled districts to determine if similar findings hold true with the caveat that state-controlled districts have the same variables that are mandated from the state.

5. A study should investigate the impact of teacher practice on student growth in middle school grades in mathematics for low-performing schools in the state of New Jersey. Within this study, there were no significant findings for teachers who taught mathematics in middle school grades.

**Conclusion**

As a nation, the United States has undergone a most significant reform in education in the last century. Almost all states are implementing or developing growth models and VAMs to better measure teacher effectiveness. As a result of this study, it would be informative to know if data pertaining to growth measures are being used to inform district leaders and inform policy. Questions to examine are: Do school leaders and teachers use the data in a formative way to inform practice and programs? If the data indicates that effective teachers are positively
correlated to student growth, what are those influences that are improving practice resulting in student growth?

Although the body of research in this area is developing and continues to increase, one can conclude, by the findings in this study, that teacher quality as represented through a teacher’s effectiveness plays a significant role in students’ academic growth. Students taught by teachers that were rated effective were likely to have typical or high growth on standardized state tests as opposed to students taught by teachers rated as ineffective. This seems like an obvious conclusion, but it is one supported by empirical evidence delineated in this study.

This study provides insight for educational leaders, researchers, and policymakers on the positive relationship between teacher practice and student growth. Consequently, the recommendation is to continue to research and explore teachers delivering instruction in their natural classroom settings to determine the other variables that influence student learning growth. The educational debates have always revolved around what is the best way to educate our children, especially in the 21st century where technology, digital learning, and global awareness are factors in teaching and learning. These debates have resulted in many educational policies, amendments, and regulations to guide the process. These spirited discussions will continue with the promise of improving education and the belief that nothing is more important than improving the teaching that occurs every day in every classroom for the benefits of students across the nation.
References


/CoventryAERA04.pdf

