

Summer 8-3-2018

# The Influence of Student/Administrator Ratios in K-6 Elementary Schools on Summative Teacher Evaluation Scores

James Ferguson Weidenborner  
*Seton Hall University*, [weidenja@gmail.com](mailto:weidenja@gmail.com)

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

 Part of the [Educational Leadership Commons](#), and the [Elementary and Middle and Secondary Education Administration Commons](#)

---

## Recommended Citation

Weidenborner, James Ferguson, "The Influence of Student/Administrator Ratios in K-6 Elementary Schools on Summative Teacher Evaluation Scores" (2018). *Seton Hall University Dissertations and Theses (ETDs)*. 2555.  
<https://scholarship.shu.edu/dissertations/2555>

THE INFLUENCE OF STUDENT/ADMINISTRATOR RATIOS IN K-6 PUBLIC SCHOOLS  
ON SUMMATIVE TEACHER EVALUATION SCORES

James Ferguson Weidenborner

Dissertation Committee

Christopher Tienken, EdD, Mentor  
Luke Stedrak, EdD  
Jaime Grinberg, PhD

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Education

Seton Hall University  
2018

© 2018 James F. Weidenborner

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

APPROVAL FOR SUCCESSFUL DEFENSE

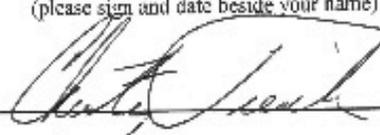
**James Weidenborner**, has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Spring Semester 2018**.

DISSERTATION COMMITTEE

(please sign and date beside your name)

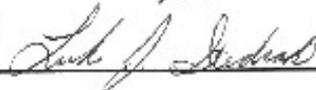
Mentor:

Dr. Christopher Tienken

 5/23/18

Committee Member:

Dr. Luke Stedrak

 5/23/18

Committee Member:

Dr. Jaime Grinberg

 5/23/18

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**

This cross-sectional, correlational, explanatory study aimed to explain the influence of administrator/student ratios on the percentage of teachers that receive effective or highly effective ratings on their summative evaluations scored by the administrators at those schools. Other school and student variables were included with administrator/student ratios as controls. The aim was to be the seminal study that provides research on factors (school and student) outside of the teacher evaluation process that influence the ratings teachers receive. Poverty and per-pupil expenditures influence teacher summative evaluation scores with regression models having  $R$  squared values of 6.3% and 6.2%. Administrator/student ratios do not influence teacher summative evaluations scored by the administrators at their schools. However, the data did provide a rationale for recommending future research on the topic.

*Keywords:* Teacher Evaluation, Tenure, Administrator/Student Ratios, TEACHNJ

## Acknowledgments

Dr. Christopher Tienken, Dr. Elaine Walker, Dr. Anthony Colella, Maureen Hoatson, and everyone at ELMP, thank you for having faith in me. You didn't forget who I am; I felt your acceptance and heard you say, "You're back, now go do it!" How lucky am I..

I believe that SHU has an absolute invaluable asset in Dr. Tienken. What I learned, and the confidence I gained, is priceless; it will empower me as I move forward in life and in my career. Thank you!

Dr. Luke Stedrak I appreciate your support greatly.

Dr. Jaime Grinberg and Dr. Katia Goldfarb: (a) Your support, crucial; (b) Your big time patience, amazing, (c) Your demonstration of love, heartwarming.

Jen Wilenta, Shannon Bellum, Christopher Cottle and all at MSU's FCST (now FSHD), please know that I would not be writing this without you. It is a privilege to be on our faculty list.

To my Gregory School colleagues. The day Carla gave me directions (no GPS back then) for my first interview marked a monumental day in my life. You have given me support and love and even some guidance (Lisa A). Dede, Stella, and Janine, your encouragement and friendship are something I am so grateful for. Lisa, Liz, Andrea, our DFs lunch club was such a cool year. Do you know even how much you mean to me!? Makeida and Sarah, thank you for support and chats... Michelle, I am heartbroken by your retirement. I have seen you stand up for your teachers, and you have certainly been there for me. My gratitude for you is immense. Thank you.

Dr. Ferguson, when you showed me how you *crushed* your problem, in a way that only you could achieve, it was pure brilliance. Believe it or not, I brag about it!

Dr. Thomas, thank you for all the years of work. Our drive to Delaware was a gift.

Dr. Lattimer, you have empowered me.

Moira, thank you so much for your caring support and enthusiasm. I wrote some of chapter II sitting on the porch at Deer Isle, warm and picturesque sunset on the horizon, one eye watching your four beautiful granddaughters dip in and out of the sound, and at peace with the world. Thank you for that.

Jeff, Marissa, Bella, and Sienna thank you for your encouragement and support. Bella and Sienna, you two are sweet, funny and very bright girls. You both ask so many great questions (that make me laugh a lot of the time), and maybe someday you smarties will be writing one of these.

Billy, Elishka, Aneska, Laura, and Lucas. We are so excited to come see you in Prague! When Trump is impeached will you consider moving back to the U.S.?

Thank you mom for your support and interest in my endeavors. It means a lot to me.

Dr. Georgia, I met you and your colleagues quite a while ago, and you have profoundly impacted my life. Dr. Kennedy, you mean the world to me. Dr. Dickstein, you helped me learn how to succeed by drilling home how I could have a “100% failure rate, that’s all, just 100%!” Joe, your friendship is such a meaningful part of my life, and I hope you know that. Lorenzo, I think downstairs in Nutley is much better than upstairs in Belleville. What do you think?!

During my coursework I had the privilege of spending time in and out the classroom with Dr. Chuck Achilles. To me he wasn’t just a master of research, he was someone that understood people. One of the great assets I have obtained from my time at Seton Hall is my ability to internalize just some of his words; the best of which came at Pals over a glass of wine and munching on a burger.

## Dedication

I dedicate my dissertation to our wonderful kitties: Bridgey and Snow Leopard. Just kidding Maggie and Lola, Daddy's "book" is 100% dedicated to you, but I do want you to know that I will always be your super corny dad.

Seriously girls, you two are the most amazing and beautiful human beings I know. Yeah, you are both incredibly beautiful on the outside, but what makes you shine is the beauty and spirit you hold on the inside. You are both individually incredible, but I want to point to some of my favorites that both of you hold... You both believe in justice... You both have passion for life and you both seek out your individual interests... You are both extremely clever... And, you both can be downright hilarious... When the three of us laugh together it's just the best thing ever (Schnitzel, Duck, Fritters—yep I went there!) I love you. I am proud of you. You make me smile.

Speaking of cheesy, there is a group of people that *always* appreciate my *incredible* sense of humor; of course, those are my students. One my favorite things is to hear of a student that has won a writing contest, a scholarship, on his/her way to a crazy career in nuclear physics or some other mind-blowing field, and even almost beating me to a doctorate. What's even more important though is the feeling I get when I see that my students have grown into happy, conscientious, mature, and moral human beings. Yay Gregory School!

Nikhil Badlani and Avery Thompson I will always miss you.

To my father. Dad, please know that I miss you, and that I can feel how proud you would be if you were here.

## Table of Contents

Abstract.....	iv
Acknowledgements.....	v
Dedication.....	vii
Table of Contents.....	viii
List of Tables.....	x
List of Figures.....	xi
<b>I INTRODUCTION.....</b>	<b>1</b>
Linking Teachers to Students.....	2
New Jersey.....	4
The Purpose of Evaluation.....	6
Instructional Supervision Beyond Evaluation.....	7
Statement of the Problem.....	8
Purpose of the Study.....	11
Research Questions.....	11
Null Hypotheses.....	12
Independent Variables.....	13
Dependent Variables.....	13
Significance of the Study.....	14
Limitations.....	15
Delimitations.....	15
Definitions of Terms.....	16
Organization of the Study.....	18
<b>II REVIEW OF LITERATURE.....</b>	<b>19</b>
Introduction.....	19
Literature Search Procedures.....	19
Methodological Issues in Studies of Predictors of Teacher Evaluation Scores.....	20
Inclusion Criteria.....	20
Review of Literature Topics.....	20
New Jersey Staff Evaluation Spreadsheet.....	21
New Jersey School Performance Report.....	21
Current Literature on Teacher Evaluation.....	22
Sputnik I.....	22
<i>A Nation at Risk</i> Standards Based Evaluation.....	24
TEACHNJ Act of 2012.....	26
Standards Based Evaluation in the TEACHNJ Era.....	30
Administrator Time.....	31
Instructional Leadership.....	33

Transformational Leadership.....	34
Theoretical Framework.....	35
Production Function.....	35
Education Production Function.....	36
<b>III METHODOLOGY.....</b>	<b>41</b>
Research Design.....	41
Research Questions.....	43
Null Hypotheses.....	43
Sample Population/Data Source .....	43
Independent Variables .....	44
Dependent Variables.....	45
Data Collection .....	45
Data Analysis .....	47
Instrumentation.....	49
<b>IV ANALYSIS OF THE DATA.....</b>	<b>53</b>
Introduction.....	53
Independent Variables.....	53
Dependent Variables.....	54
Procedure.....	55
Results.....	56
Correlations.....	61
Simultaneous Multiple Regressions.....	63
<b>V CONCLUSIONS AND RECOMMENDATIONS.....</b>	<b>70</b>
Conclusions.....	70
Recommendations for Policy.....	73
Recommendations for Practice.....	78
Recommendations for Future Research.....	82
<b>REFERENCES.....</b>	<b>85</b>

## List of Tables

Table 1.	Example of the Difference in Total Observations Prior to and After TEACHNJ.....	10
Table 2.	Data Retrieved from the NJDOE and Utilized in the Excel Spreadsheet.....	46
Table 3.	Simultaneous Regression Model.....	49
Table 4.	Variables and Names of Independent Variables.....	54
Table 5.	Descriptive Statistics Table.....	57
Table 6.	Descriptive Statistics for Teachers Receiving Effective Scores on Summative Evaluations.....	58
Table 7.	Effective Evaluation Scores Normality Table.....	59
Table 8.	Descriptive Statistics for Teachers Receiving Highly Effective Scores on Summative Evaluations .....	60
Table 9.	Effective Evaluation Scores Normality Table.....	61
Table 10.	Pearson Correlations of the School and Student Independent.....	62
Table 11.	Pearson Correlations of the School and Student Variables. and Effective and Highly Effective Percentages.....	63
Table 12.	Percentage of Teachers Scored Effective Model Summary.....	64
Table 13.	Percentage of Teachers Scored Effective ANOVA Table.....	64
Table 14.	Effective Percentage Coefficients Table.....	65
Table 15.	Percentage of Teachers Scored Highly Effective Model Summary.....	66
Table 16.	Percentage of Teachers Scored Highly Effective ANOVA Table.....	66
Table 17.	Highly Effective Percentage Coefficients Table.....	67

## List of Figures

Figure 1. Clinical Supervision.....	23
Figure 2. Simultaneous Regression Framework.....	47

## CHAPTER I

### INTRODUCTION

The No Child Left Behind Act of 2001 (NCLB, 2002) brought teacher quality to the forefront of public school education policy. The bill required that teachers meet the newly created status of *highly qualified teacher*. For elementary school teachers, highly qualified status meant that they had to hold at least a bachelor's degree and have passed their respective state teacher certification examinations. Secondary teachers, or any teacher teaching a "core" subject, defined as math, English language arts, social studies, science, and so forth had to be certified in the specific subject that they taught.

The law added another layer of accountability for quality teaching: the use of results from standardized testing for all students in Grades 3–8 and once in high school, to measure the impact of the teacher on student learning. Moreover, student standardized test results had to show Adequate Yearly Progress (AYP), a specified improvement in student test scores on a yearly basis. Lastly, the law included a clause stating that by 2014, 100% of the country's public school students were expected to achieve proficiency on their state's standardized tests (NCLB, 2002). NCLB lasted through the Bush administration and into most of the Obama presidential administration.

The Obama administration instituted the Race to the Top competitive grant program (RTTT). The program offered federal financial incentives for states that evaluated teachers and school administrators based, totally or in part, on student results from state-mandated standardized tests (Baker, Oluwole, & Green, 2013).

The former United States Secretary of Education, Arne Duncan, initiated an NCLB waiver program during the Obama administration whereby states were required to use standardized tests

in the evaluation of teachers and school administrators. Evaluations that encompassed measuring test scores did not cease when the Every Student Succeeds Act (ESSA) of 2015 eventually replaced NCLB (Klein, 2015). The Federal Department of Education gave states the choice to continue the practice of evaluating teachers and school administrators using test results, and as of 2017, approximately 40 states still continue the practice.

### **Linking Teachers to Students**

According to Baker et al. (2013), there are two popular methods used to link teachers to their students' test scores: (a) value-added measure (VAM), which is also referred to as value-added assessment (VAA) and (b) student growth percentile (SGP).

In 1992, education officials in Tennessee were the first to use VAAs as a tool for educational reform (Jordan, Mendro, & Wesinghe, 1997; Sanders & Horn, 1998), although VAAs were not widely used until after NCLB was signed. The practice of using VAAs to evaluate and identify teachers' proficiency levels was flawed from the start, as its creator, William Sanders, failed to acknowledge the influence that student socioeconomic status (SES) and class-size reduction (CSR) each had on student achievement (Sanders & Horn, 1998). Evidence to the former has been available since 1966 with the Coleman Report (Coleman et al., 1966) and the latter from the Tennessee STAR study data, initiated in 1985 (Achilles, 2012; Finn & Achilles, 1990; Mosteller, 1995; Tienken & Achilles, 2006). Nonetheless, VAA use started the process of connecting teacher ratings to student test scores. Yen (2007) wrote that VAAs lacked in demonstrating causality and could only be used to describe the growth (or lack of) in student achievement but not a teacher's impact on test scores.

Starting in the summer of 2008, the officials from the Colorado Department of Education (CDE) started using an SGP model that was called a student growth model, and more states

followed using SGP models for evaluations in the years thereafter (Betebenner, 2009). In 2011 Betebenner cited Braun (2005), Rubin, Stuart, and Zanutto (2004), Ballou, Sanders, and Wright, (2004), and Raudenbush (2004) as sources stating:

The primary thrust of growth analyses over the last decade has been to determine, using sophisticated statistical techniques, the amount of student progress/growth that can be justifiably attributed to the school or teacher—that is, to disentangle current aggregate level achievement from effectiveness. (p. 1)

Betebenner (2011) explained that an SGP is a norm-referenced tool to measure typical student achievement growth relative to his or her academic peers. For SGP measurement, an academic peer group consists of students that scored similarly on the previous year(s) achievement tests. Even though criterion referenced tests are used, a norm-referenced value is created. After an achievement peer group is established, then that cohort of students are measured against each other based on the scores they received on their current year's test and are assigned a percentile score (i.e., a student with a percentile score of 65% scored higher than 65% of the students in his or her peer group with similar scores on previous year's test). An SGP value for a teacher is simply the average percentile of all of his or her students within their individual peer groups (Castellano & Ho, 2013); that is, a teacher is given a percentile score between 1 and 99.

Ehlert, Koedel, Parsons, and Podgursky (2013) stated that SGPs were never intended to be used to measure teacher effectiveness. Betebenner (2009) explained the process thoroughly but also noted that the results obtained are descriptive and not causal. Baker, Barton, and Darling-Hammond, Haertel, Ladd, Linn, Ravitch, et al. (2010) found that SGPs became popular when VAMs were found not to be connected to teacher effectiveness.

Because of the claim that standardized test scores represent an objective measure of student academic achievement, an evaluation system that encompasses VAAs or SGPs must include a quantifiable rubric where teachers are placed into categories of effectiveness, regardless of the percent of the actual summative evaluation that is attributed to a VAA or SGP. The scale used by the Colorado Department of Education is an example: *basic, partially proficient, proficient (meets state standard), accomplished, or exemplary* (CDE, n.d.). State departments of education across the U.S. adopted similar “categories” of effectiveness that teachers are placed into.

### **New Jersey**

Officials at the New Jersey Department of Education (NJDOE) began experimenting with SGPs to describe teacher effectiveness and student academic growth in 2010. Officials started with 2 years of training and pilot programs in various districts (NJDOE, n.d.a). SGP evaluations became law in New Jersey public education when Governor Chris Christie signed the TEACHNJ Act of 2012 (and the ACHIEVENJ regulation bill that went along with it) into law (Ch. 26; C.18A:6-117). More piloting and experimenting was undertaken, and as of 2017-2018, New Jersey is in the fifth year of using an evaluation system that incorporated SGP use. The four performance categories teachers and administrators are placed in are: (a) ineffective, (b) partially effective, (c) effective, or (d) highly effective.

NJDOE officials approved 18 different teacher evaluation models for adoption by school districts (NJDOE, 2015b). As of 2013, 60% of reporting school districts (291 districts) used Charlotte Danielson’s, *A Framework for Teaching* (Mooney, 2013), which got its start in 1996 and has since evolved into numerous updated editions (Danielson, 2015). In this model for professional development, now also used for teacher evaluation (Danielson, 2010; Evans, Willis,

& Moretti, 2015), there are four domains, 22 component indicators, and 76 smaller elements of professional practice (Danielson, 2007).

Each approved evaluation model must include the four performance category ratings for teachers. Three of the other most commonly used evaluation models in New Jersey are the Stronge Teacher and Leader and Effectiveness Performance System, the Teach for Education and Learning (McREL) Teacher Evaluation Standards, and Marzano's Causal Teacher Evaluation model. The number of districts that adopted each model was 53, 45, and 44 respectively (Mooney, 2013). Regardless of the model, officials in each school district take the scores from their evaluation tools, combine them with SGP scores (calculated by the NJDOE) and Student Growth Objective (SGO) scores to arrive at a summative teacher evaluation rating. An SGO is a non-scientific measure in which a teacher creates a pretest for a group or groups of student(s), as a baseline to measure predicted growth on a posttest. Then a score is assigned to the teacher based on how his or her students scored on the posttest that corresponds to the four-tiered evaluation rubric outlined in TEACHNJ. Teachers can be identified as (a) ineffective, (b) partially effective, (c) effective, or (d) highly effective based on an arrived upon score. According to the TEACHNJ Act, there are four scores that teachers can receive on their summative evaluation in NJ: (a) highly effective, (b) effective, (c) partially effective, or (d) ineffective. The TEACHNJ Act requires that every school must report the total number of teachers that have received a score in each of the four possible summative evaluation scores. However, it is also stated that every school must have a category total blacked out if the number of teachers receiving that score is less than 10. Not a single school in NJ has a reported 10 or more partially effective teachers (NJDOE, n.d.f).

Although SGPs are a part of the formal teacher evaluation process statewide, the use of SGPs are now almost negligible, in that they are only 10% of a teacher’s summative evaluation score (NJDOE, 2015a). Teacher practice, “measured according to the district-chosen observation model” (NJDOE, n.d.b, slide 4), makes up the largest share of a teacher’s summative evaluation score in New Jersey, as per the four-tiered scoring system.

The TEACHNJ act of 2012 mandated specific parameters for how teacher practice is evaluated. School administrators must conduct at least three observations on every teacher under their supervision, including tenured teachers. Prior to the passage of the 2012 law, non-tenured teachers were required to receive at least three classroom teaching evaluation observations per year and tenured received at least one mandatory observation. The law also extended the time it takes for a new teacher to obtain tenured status from 3 years to 4 years. In addition to the now scant requirement for SGPs the law also required a SGO to be used; there are multiple steps and deadlines that must be completed hand in hand with administrators (NJDE, n.d.c). Paramount to all of the provisions of TEACHNJ is the clear purpose and goal of this law was to have a standardized system to evaluate teachers and “*streamlining the tenure arbitration system for all staff members*” (NJDE, n.d.d, p. 4).

### **The Purpose of Evaluation**

In section 18A: 6-118 of TEACHNJ the purpose (or 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” (para. a). Farhat (2016) stated that teacher evaluation is not only supposed improve classroom instruction but also promote the professional development of faculty (Farhat, 2016). Darling-Hammond, Wise, and Pease (1983) wrote that there is a “need for context-

specific strategies for improving teaching rather than system-wide hierarchical efforts.” (p. 311). And, Lewis (1982) offered that “no single model [of instruction] will result in effective learning...” and that ‘an evaluation system must respect the uniqueness of each individual staff member’ (as cited in Darling-Hammond et al., 1983, p. 311).

Teacher evaluation systems that rely on checklists and a predetermined set of behaviors have been found to be of little value (Peterson, 2004; Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984). Evaluation systems that treat teachers as inanimate numbers or scores and do not consider them as valuable individuals have little credibility (Wise & Darling-Hammond, 1984-1985). Part of the stated purpose of TEACHNJ is to improve classroom instruction and promote professional development, which is aligned with the findings of the extensive literature review conducted by Darling-Hammond et al. (1983). However, having the purpose of evaluation as the basis for personnel decisions, as in TEACHNJ, was not found in the Darling-Hammond et al. (1983) review. Furthermore, the information put out by the state that interprets TEACHNJ to the public (NJDOE, n.d.a; NJDOE, n.d.b; NJDOE, n.d.d; NJDE, n.d.e) paints teacher evaluation as the means to make teacher personnel and tenure decisions.

### **Instructional Supervision Beyond Evaluation**

Oliva and Pawlas (2004) asserted that the role of the supervisor extends beyond simply improving instruction; administrators must also “increase satisfaction, create learning communities, expand student understanding in the classroom, and analyze cultural and linguistic patterns in the classroom” (p. 5). Principals are charged with creating and maintaining budgets and a plethora of administrator actions that are outside the realm of teacher evaluation (Furney, Aiken, Hasazi & Clarke-Keefe, 2005; Pogodzinski, 2013). Stronge, Holly, and Catano (2008) dedicated eight chapters to “What it Means to Be an Effective Principal” (p. vii), which includes

an excessive amount of qualities, behaviors, and specific examples of what effective principals undertake; however, just *one* chapter was dedicated to teacher evaluation.

With all the responsibilities effective administrators must take on, time and stress management are crucial factors (Grissom, Loeb, & Mitani, 2014). Stated in the TEACHNJ Guide, “At its core TEACHNJ reforms the process of earning and maintaining tenure” (NJDOE, n.d.d, p. 1). For administrators to accomplish what is at the crux of TEACHNJ, the time they dedicate to teacher evaluations must trump all.

### **Statement of the Problem**

The NCLB Act of 2001 called for teachers, principals, school districts, and entire state public education systems to be held accountable for student test score results, using school and district mean scores and a calculated AYP. The use of AYP led to an increase in the use of VAA, which, in turn, led further to the creation and use of the SGP. In order to use SGP and VAA more objectively, quantifiable rating scales had to be created. Therefore, in some cases, teachers were evaluated based on rubrics of teacher practice that were not meant to be used for evaluation, and the scores from those rubrics became one data point for the use of SGP and VAA. The SGP and VAA resulted in teacher practice categorized as (a) ineffective, (b) partially effective, (c) effective, or (d) highly effective based on the SGP or VAA score.

“In June 2010, the NJ Legislature adopted Governor Chris Christie's proposed FY11 budget, cutting over \$1.1 billion, or almost 15%, in state aid” (Education Law Center, n.d., para. 6). Bret Schundler, NJ State Commissioner of Education, explained the cuts in a March 19, 2010, letter to Chief School Administrators and School Business Administrators. It was on the heels of these significant cuts that the TEACHNJ Act of 2012 was passed, forcing school administrators in the state to use rubric-based evaluation instruments; many of which were not created for

teacher evaluation or use with SGP or VAA. Although much attention was given to the use of SGP in teacher evaluation, a major component of the law was a substantial increase in the observations that administrators were required to perform. K–5 principals now had to compute SGPs for all of their fourth and fifth grade teachers as well as conduct a substantial increase in observations for all of their teachers, tenured and non-tenured. The specific changes made were:

- According to “Requirements and Resources for Teacher Evaluations in 2014-15” put out by NJDOE (n.d.g), ALL teachers, including tenured teachers, had to be observed three times a year (however, starting in the 2016-2017 school year it was lowered to two total for tenured teachers).
- One observation had to be conducted by multiple administrators simultaneously in collaboration with each other for all non-tenured teachers.
- Tenure status was not to be granted to teachers until the completion of their *fourth* year teaching (as opposed to the third year before).

In addition, school administrators were obligated to work with teachers to create non-scientific pretest-posttest measures known as SGO, the results of which were included in summative evaluations. In a 2012 report on the state education budget Christopher Cerf, former Acting Commissioner of Education, wrote, “The research could not be clearer that great teachers are more important to learning outcomes than class size” (p. 28). Essentially, faulty research was being used to justify the argument that cutting education spending would benefit students across the board, including those in districts containing large populations of children of low socioeconomic status. In this report Cerf also laid the blame for achievement deficiencies on “ineffective” teachers.

Table 1 is an example I created to demonstrate the amount of observations an administrator would have to perform before and after the passing of TEACHNJ. I used 20 teachers: 16 tenured, two non-tenured in their first year, and two non-tenured in their fourth year.

Table 1

*Example of the Difference in Total Observations Prior to and After TEACHNJ*

Prior to TEACHNJ:	Total observations through 4 years	After TEACHNJ	Total observations through 4 years
16 tenured teachers: 16 x 1	16 observations	16 tenured teachers: 16 x 3	48 observations
Two teachers at year one: 2 x 3	6 observations	Two teachers at year one: 2 x 3	6 observations
Two teachers at year four: 2 x 1	2 observations	Two teachers at year four: 2 x 3	6 observations
Total	24 observations	Total	60 observations

Administrators observe, evaluate, and determine the effectiveness of teachers. Ergo, the daunting task of realizing the ultimate goal of TEACHNJ was placed on the laps of administrators. The numerous other important tasks and behaviors described by Furney et al. (2005), Hoy, Tarter, and Woolfolk Hoy (2006), Hoppey and McLesky (2013), and others were seemingly ignored. Little quantitative evidence exists on whether the amount of available time an administrator has, which diminishes with every additional teacher to be evaluated, is a factor that influences the final summative evaluation ratings of teachers in New Jersey.

## **Purpose of the Study**

My purpose for this study was to explain the influence that administrator/student ratios had on the summative scores that teachers received from their evaluating administrators. In addition, I aimed to explain the amount of variance in summative teacher ratings accounted for by administrator/student ratios when controlling for other student- and school-level factors that influence the available time administrators have to dedicate to conducting evaluations.

The correlational, explanatory, cross-sectional design (Johnson, 2001) used was based in the production function theory (which is most often used to relate various variables to student achievement) can also be used for other outcomes (Hanushek, 2010). In general terms, production function theory is “a function that relates various inputs to education including those of families, peers, and schools to the maximum level of student achievement that can be obtained” (Hanushek, 2010, p. 407).

Looking beyond the TEACHNJ Act of 2012, the aim of this study is to provide policy makers, from the state level down to local boards of education, data that can be utilized to (a) look into adding administrator/student ratios as a part the of evolving equation of how to improve educational practices, (b) direct resources, and (c) at the state level re-examine the weight given to teacher evaluation results in teacher tenure decisions

## **Research Questions**

There are four ratings that teachers can receive on their summative evaluation in New Jersey: (a) highly effective, (b) effective, (c) partially effective, or (d) ineffective. For this study, partially effective and ineffective have been combined into one category because not a single school in the state has identified at least 10 teachers as ineffective, and by law a district may not publish the numbers for any of the four categories if the total number of teachers that received

that score is less than 10. The overarching research question that guided this study was: What is the influence of administrator/student ratios on teacher rating scores, measured by percentages of teachers receiving effective or highly effective by administrators at their schools, when controlling for different student and school variables?

**Research Question 1:** What is the influence of school administrator/student ratios at elementary schools containing grades PreK–6 (as identified by containing three consecutive grades or more, with no class above the sixth grade) on the percentage of teachers being rated effective/highly effective on their summative evaluations scored by the administrators at those schools?

**Research Question 2:** What is the influence of school administrator/student ratios at elementary schools containing grades PreK–6 (as identified by containing three consecutive grades or more, with no school class above the sixth grade) on percentage of teachers being rated highly effective on their summative evaluations scored by the administrators at those schools?

### **Null Hypotheses**

**Null Hypothesis 1:** No statistically significant relationship exists between school administrator/student ratios and the percentage of teachers scored effective/highly effective by administrators at those schools.

**Null Hypothesis 2:** No statistically significant relationship exists between school administrator/student ratios and the percentage of teachers scored highly effective by administrators at those schools.

## **Independent Variables**

In this study the independent variables were taken from the 2015-2016 NJ School Performance Report and The Taxpayers Guide to Education Spending 2016 that was found on the New Jersey Department of Education website. The variables included were administrator/student ratios as well as specific variables categorized by student and school.

**Student variables:** Percentage of students that were English Language Learners (ELL), percentage of students qualifying for free or reduced lunch, and percentage of students with disabilities.

**School level variables:** Administrator/student ratios and per pupil expenditure.

## **Dependent Variables**

The dependent variables were taken from the Staff Evaluations 2015-2016 spreadsheet, found under the heading Major DOE Reports on the Data tab of New Jersey Department of Education website. The spreadsheet contains the reported scores from the evaluation rubric outlined in the TEACHNJ Act of 2012, created to be the standardized statewide teacher evaluation tool. The four tiers encompassed in the rubric are (a) highly effective, (b) effective, (c) partially effective, or (d) ineffective.

Teachers that receive a partially effective or ineffective score are to be put on a corrective action plan (CAP) that consists of various improvement goals. If a teacher is scored ineffective, then in the following year his or her superintendent must file a charge of “inefficiency” (otherwise known as tenure charges), and if scored partially effective their superintendent may defer charges for a year by writing the state board of education a letter that explains the warranting extraordinary circumstances. If a teacher does not make the effective threshold in 3 straight years, a superintendent must file a charge of inefficiency (NJDOE, n.d.d, p. 5).

## Significance of the Study

Only PreK–6 schools that contain three consecutive grades or more, with no class above the sixth grade, were used in this study. NCLB identifies elementary school teachers to be *highly qualified* if they hold an elementary school teaching certificate, whereas secondary teachers, must have a content area teaching certificate in order to be considered *highly qualified* (NCLB, 2002). Elementary school teachers are a streamlined sample and are less likely to be evaluated based on their level of content knowledge (which can vary greatly) but more so on their teaching. In addition, as found in a meta-analysis conducted by Greenlee (2007), secondary schools allocate budgets to “miscellaneous categories” (p. 241); however, elementary school budgets are directed towards classroom instruction and curricula. Elementary schools consisting of grade levels that run from primary to upper elementary (e.g., K–5) provide the cleanest sample of faculty for conducting analyses on teacher evaluations. There is a dearth in research that measures the influence of administrator actions—outside of the provisions of TEACHNJ—that influence teacher evaluation scores.

Leech, Barret, and Morgan (2011) stated, “It is preferable to use the hierarchical method when one has an idea about the order in which one wants to enter predictors and wants to know how predictions by certain variables improves on prediction by others” (as cited in Sammarone, 2014, p. 14). This study has used administrator/student ratios as well as other student and school variables in a hierarchical regression to predict teacher evaluation scores.

This study can serve as catalyst to drive further research in discovering factors that influence teacher evaluation scores. TEACHNJ is a law that’s primary purpose was, “*streamlining the tenure arbitration system for all staff members*” (NJDE, n.d.d, p. 4), but if factors outside of the provisions of the law itself predict how teachers are scored, then there is a

significant problem.

### **Limitations**

This study was a correlational, explanatory, cross-sectional design using quantitative methods (Johnson, 2001). The inability to express a cause and effect relationship is a limitation to all non-experimental research. This explanatory cross-sectional study only used reported teacher ratings from the 2015-2016 school year; therefore, there are limitations based on alternative relationships that could be found based on the year of the data used.

This study was designed under the notion that teacher summative evaluation scores are in the sole control of assigned evaluating administrators. The only exception being that 10% of the evaluation scores for teachers in Grades 4, 5, and 6 comes from their SGP score (which is produced by the NJDOE using student test scores). For the 2015-2016 school year SGPs accounted for 10% of summative evaluation scores for fourth, fifth, and sixth grade teachers.

SGOs comprise 20% of all teachers' summative evaluation scores. SGOs are data-based pretest/posttests that are worked on collaboratively between a teacher and his or her evaluating administrator. These tests are not in the sole control of evaluating administrators.

Educational level of the faculty (percentage of teachers' holding a master's degree or more) was not taken into account. Only PreK-6 schools were used in this study, and secondary subject-specific degrees and certifications are not needed for elementary school faculty.

### **Delimitations**

Data were taken from PreK-6 schools that contain three consecutive grades or more with no class above the sixth grade. Data from elementary schools that only contained two grades (e.g., K-1) or stand-alone single grade schools were not used. The data were limited to only public schools; parochial, charter, and private schools were not considered.

If highly effective and effective numbers are reported, then all the other scores logically have to be a combination of partially effective and ineffective, as the school total number of teachers evaluated is reported. However, if only effective, or only effective and partially effective scores are reported, one has no way of precisely identifying how many teachers have been scored ineffective or highly effective. Therefore, only schools that have numbers reported for both effective AND highly effective were used.

### **Definitions of Terms**

The terms defined below were retrieved from The State of New Jersey Department of Education's website.

**Achievement gap** – Is the variance of student achievement between groups.

**Administrator/Student ratio**- Is the number of students for every one administrator in a public school.

**Class-Size reduction (CSR)**- Is the process of reducing the number of children in a classroom. CSR is not related to pupil–teacher ratio (PTR).

**District Factor Group (DFG)** classifications are based on U.S. Census data and are revised every 10 years. The DOE uses DFG data to analyze the relationship between student achievement and the socioeconomic status of the communities in which they reside. The six census data indices used in the DFG statistical model include the percentage of each district's population with no high school diploma, the percentage with some college education, and the poverty level and unemployment rate of the district, as well as the residents' occupations and income. The analysis and weighting of these components is used to produce a statistical score for each district, which is then ranked and placed into one of eight groupings —A, B, CD, DE, FG,

GH, I, and J. Each grouping consists of districts with similar factor scores. I and J districts score highest on the socioeconomic scale.

**Limited English Proficient (LEP) students-** This is the percentage of LEP students in the school. It is calculated by dividing the total number of students who are in Limited English Proficient programs by the total enrollment.

**No Child Left Behind -** The No Child Left Behind Act (NCLB, 2002) of 2001 was signed into law on January 8, 2002, by President Bush. The Act represents the president's education reform plan and contains the most sweeping changes to the Elementary and Secondary Education Act (ESEA) since it was enacted in 1965. NCLB changes the federal government's role in K–12 education by focusing on school success as measured by student achievement. The Act also contains the president's four basic education reform principles of stronger accountability for results, increased flexibility and local control, expanded options for parents, and an emphasis on teaching methods that have been proven to work.

**Per-Pupil expenditure-** A school district's total budget divided by the total number of students in that district.

**Students with disabilities -** This is the percentage of students with an Individualized Education Program (IEP), including speech, regardless of placement and programs. This is calculated by dividing the total number of students with IEPs by the total enrollment.

**Teacher Effectiveness and Accountability for the Children of New Jersey (TEACHNJ Act of 2012) -**This is an act concerning school employees, revising various parts of the statutory law, and supplementing Chapters 6 and 28 of Title 18A of the New Jersey Statutes.

## **Organization of the Study**

Chapter I provided informational background and set forth an overview of the problem related to teacher summative evaluations scores on the New Jersey teacher rating rubric outlined in the TEACHNJ Act of 2012 and the relationship to administrator/student ratios. The evaluation systems and state rubric are intended to inform tenure decisions statewide; however, the process of conducting evaluations is time consuming and, in turn, replaces time that could be dedicated to other actions that have been deemed as effective practices in past research. I sought to determine variance in teacher evaluation ratings scored by their administrator that was accounted for by administrator/student ratios while controlling for other variables, such as per-pupil expenditures.

Chapter II included a review of the literature on the identified student- and school-level factors as well as the theoretical framework related to administrator/student ratios and teacher summative evaluation ratings scored by their assigned administrators.

Chapter III, in tandem with Chapter I, explained the design methods and procedures for this study. The data used were collected from the Staff Evaluations 2015-2016 spreadsheet (NJDOE, n.d.e) and the New Jersey Performance Report for the year 2015-2016 (NJDOE, n.d.f), both found on the New Jersey Department of Education website.

Chapter IV presented the data and statistical findings of the study.

Chapter V specified a statistical summary and data implications for administrative practices and policies. Detailed recommendations and conclusions derived from the research findings were presented, and suggestions were made for future research.

## CHAPTER II

### REVIEW OF THE LITERATURE

#### **Introduction**

The purpose of this correlational, cross-sectional, study was to explain the influence that administrator/student ratios had on the summative scores that teachers receive from their evaluating administrators. Based on the overarching research question, the literature review was composed of the following sections: NJ Staff Evaluation Report, NJ Performance Report, teacher evaluation, teacher evaluation in NJ, student variables, school variables, and administrator time.

The review functioned to demonstrate empirical, descriptive, and theoretical literature that attempted to explain the influence that administrator variables, school variables, and student variables have on teacher evaluation scores in order to inform school leaders, board of education members, bureaucrats, and researchers. Specific attention was paid to the extensive research on teacher evaluation and administrator time.

#### **Literature Search Procedures**

As outlined by Boote and Beile (2005), I used multiple databases for my search, such as ProQuest Education, ProQuest Central, ERIC, and Education Research Complete; using both the Seton Hall University Library and the Harry Sprague Library at Montclair State University; and after using databases I also searched further through specific peer-reviewed journals. All variables and related terms were searched by keyword: for example, *teacher evaluation*, *successful administrators*, and *per-pupil expenditure*. The reviewed literature included experimental, quasi-experimental, and meta-analysis.

## **Methodological Issues in Studies of Predictors of Teacher Evaluation Scores**

There exists a dearth of studies that investigate the presence of connections between administrator/student ratios and reported teacher evaluation scores in New Jersey. In addition, no research was found that relates administrator/student ratios to teacher evaluation or effectiveness countrywide. Furthermore, there is a lack in research of any kind that predicts the influence of school or student factors (other than test scores) on teacher evaluation rating percentages, scored by their evaluating administrators. Weisberg, Sexton, Mulhern, and Keeling (2009) in a study using approximately 15,000 teachers, and 1,300 administrators across 12 school districts in four states found, “All teachers are graded good or great” (p. 10). This descriptive analysis was the only data that could be found.

### **Inclusion Criteria**

Research used in this review had to meet at least one of the following criteria in order to be included:

1. studies which were experimental, quasi-experimental, as well as non-experimental with control groups;
2. peer-reviewed research including dissertations and government reports;
3. published within the last 25 years;
4. studies that included teacher evaluation;
5. studies that focused standardized teacher evaluation;
6. any literature found in a government report that meets the above criteria; and
7. seminal works.

### **Review of Literature Topics**

I reviewed literature on administrator time, the evolution of teacher evaluation, and leadership styles, all topics that can be influenced by the variables found in the NJ School Performance Report. These variables (such as administrator/student ratios), in turn, influence how administrators score summative evaluations for the teachers at their schools. Outlined in the

TEACHNJ act of 2012 teachers can be scored and identified as: (a) ineffective, (b) partially effective, (c) effective, or (d) highly effective.

### **New Jersey Staff Evaluation Spreadsheet**

Reports of teacher evaluation scores are made available at the individual school and district level for all public schools. For each *effective* category, “Records that have *n*-size < 10 are suppressed” (NJDE, n.d.e, para. 2) in order to insure that individual teachers cannot be identified by the public. In the 2014-2015 Staff Evaluation report not a single school reported an *n*-size  $\geq 10$  for the category of *ineffective*, and at the district level only Newark and Camden reported actual *ineffective* numbers (90 and 69 respectively). In the entire state only 169 teachers out of 106,542 were scored *ineffective*, which means that only 10 teachers outside of Newark and Camden were scored *ineffective* statewide.

For the categories of *partially effective*, *effective*, and *highly effective* it varied from school to school and district whether there was an *n*-size  $\geq 10$ . My research examined school-level reported scores at K–5 elementary schools that reported scores for the *effective and highly effective* (with an *n*-size  $\geq 10$ ) categories.

### **New Jersey School Performance Report**

The stated focus of the Performance Report is to offer school- and district-level information concerning college and career readiness and highlight data that NJDOE deems related to student outcomes (NJDOE, n.d.e). The Performance Report spreadsheet contains distinguishing information including achievement scores, descriptive student and staff data, and school climate. My research incorporated student and school variables.

## Current Literature on Teacher Evaluation

Teacher evaluation, throughout most of our history, has been practiced with the intent of improving teacher performance and, ultimately, increasing student learning. “Despite what appears to be a concerted effort across the past several decades, teacher evaluation did not work the way it is intended” (Xu, Grant, & Ward, 2016, pp. 203–204).

### Sputnik I

Public education in the U.S. hit the crisis mode when the Russians successfully launched Sputnik I. Powell (2017) wrote that even though the satellite was rather simple, it embarrassed America, and education had become a national defense issue. In 1958, millions of dollars were invested in the structural reform of American public education when Congress enacted the National Defense Education Act (NDEA). New curricula were developed and evaluations created in attempt to cure the phenomena (Hogan, 2007). The fear brought on compulsory evaluation of American public education (including individual teachers) and the assessment process has continued to evolve ever since.

The nationwide reaction to Sputnik I resulted in the widespread use of Robert Goldhammer’s *Clinical Supervision* a decade later. Written in 1968, and published after his death in 1969, it was not truly actualized until the publication of Morris Cogan’s 1973 book titled *Clinical Supervision*. There are five basic components to clinical supervision: (a) pre-conference, (b) observation, (c) analysis and strategy, (d) supervision conference, and (e) post-conference strategy. Since Cogan’s seminal work, many have adopted and tweaked the format without changing it. Glickman, Gordon, and Ross-Gordon (2004) created the following sequence:

**Step 1:** Preconference with the teacher where issues discussed are: reason, focus, method, and

time of the observation.

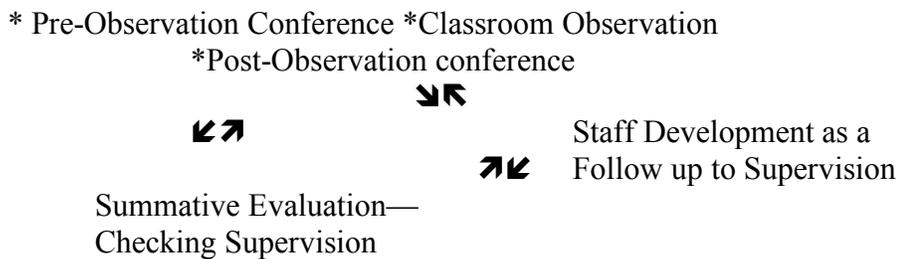
**Step 2:** Observation of classroom where the teacher is observed performing in the natural classroom setting and description of occurrences are recorded in an observation instrument.

**Step 3:** Analysis and interpretation of the data collected during the observation takes place.

**Step 4:** Post-conference when the supervisor meets with the teacher to discuss the analysis of the observation and finally “produce a plan for instructional improvement.” Objectives are set for the teacher to meet for the following observation.

**Step 5:** The critique of the supervisory steps by the teacher to give the supervisor implications on what was valuable in the process and what needs modifications and improvement.

Zepeda and Mayers (2014) specified clinical supervision as an indefinite process to be worked on between teacher and administrator (p. 19):



*Figure 1: Clinical Supervision*

Within the clinical supervision model, Madeline Hunter introduced a component model for effective teaching, which was quickly and widely adopted as criteria for teacher evaluation. The components were: (a) anticipatory set, (b) stated objective, (c) teacher input, (d) modeling, (e) checking for understanding, (f) guided practice, and (g) independent practice (Hunter, 1984). Hunter (1980) also influenced supervisory conferences and identified 6 types of supervisor/teacher conferences, with the first five being various instructional conferences and the

last being a summative evaluation. While her seven-step model of teaching could be considered behaviorally rigid, with clinical supervision dominating the process for evaluating a teacher's instructional skills, several studies helped define evaluation even further. Darling-Hammond et al. (1983) conducted an extensive review of the current teacher evaluation literature and came to the conclusion that effective evaluation consists of providing information and purposeful improvement at the individual (teachers and staff) and organizational levels (school and district). They discussed the use of standardized definitions in teacher evaluation but ultimately found, "Research on teacher performance and teaching effectiveness does not lead to a stable list of measurable teaching behaviors effective in all teaching contexts" (p. 320).

Cawelti (1982) wrote that evaluation should provide for a united school vision under which teachers use the process as a tool for their individual professional growth, and McGreal (1983) asserted that evaluation should provide formative improvement and summative accountability. In an expansive study of four distinctly diverse school districts—derived from an initial survey of 32 school districts using criteria such as method and process of evaluation, purpose of evaluation, organizational structure of the district, and so forth—Wise et al. (1984) studied organizational commitment, evaluator competence, teacher–administrator collaboration, and strategic compatibility. They concluded that evaluation systems: (a) must be connected to community goals; (b) commitment to evaluation has to outweigh the use of checklists; (c) the purpose and process of evaluation need to be coherently matched; (d) resources and political support ought to be perceived to have utility, used with efficiency; and (e) teachers are stakeholders that are responsible for evaluation improvement.

### ***A Nation at Risk* and Standards-Based Evaluation**

Twenty-six years after the National Commission on Excellence in Education (1983)

published *A Nation at Risk: The Imperative for Educational Reform* asserting that public schools in America were in need of serious reform once again, the authors wrote, “History is not kind to idlers” (p. 6). Our country’s shortcomings on the shoulders of educational accountability and a subsequent consequence of this call to action was that the improvement of teacher accountability through evaluation was imperative countrywide. Gage and Needels (1989) posited that the focus on teacher evaluation systems was a logical step. According to Darling-Hammond (1990), at this point it was clear that a standardized system of accountability was needed and that our evaluation systems needed to adapt and incorporate specific streamlined criteria (McLaughlin & Pfeifer, 1988).

Less than 20 years later, teacher accountability was forced nationwide when the federal government passed the NCLB Act of 2001 that “asks the states to set standards for student performance and teacher quality” (Paige, 2004, p. ii). Teachers had to meet the threshold of *highly qualified*; however, Toch and Rothman (2008) noted that the term only identified teachers’ qualifications and not their performance as teachers. In 2009, with the passing of the American Reinvestment and Recovery Act (ARRA), and the 4.35 billion-dollar Race to the Top (RTT) grant program found within, teacher accountability became the giant carrot at the finish line that states were forced to chase. A component of the law was to reward and retain highly effective teachers (U.S. Department of Education, 2009). “RTT encouraged states and districts not only to revamp their teacher and principal evaluation policies, but also to use evaluation results to make personnel decisions” (Aldeman, 2017, p. 62). Tangential to rewarding schools and school districts when evaluation is connected to student test scores was the use of merit pay systems, where teacher compensation (not just evaluation) is in some way connected to students’ test scores. Belfield and Heywood (2008) posited that merit pay decreases job satisfaction in

teachers because of an increased workload and the relative level of randomization in actual student test score results. In addition, Gius (2014) found no difference in teacher job satisfaction between those that received merit pay and those that did not.

Using student test scores as a part of the process of measuring school success continued, and teacher evaluation became a major reform tool to improve instruction and raise achievement levels (Goldhaber, 2015); standards-based evaluation became a necessity. Milanowski and Heneman (2002) stated three components to standards-based evaluation: (a) alignment to current literature that identifies strong teaching, (b) specific multi-level standards that are a reflection of high performing teachers, and (c) the use of multiple data sources. The additional source of data was linking teacher performance to student test scores, which began with VAAs (also known as Value Added Measurement [VAM]). These analyses were first used in Tennessee in 1992 (Jordan, Mendro, & Weerasinghe, 1998; Sanders & Horn, 1998). Referring to the wide spread use of standards based reform, Xu, Grant, and Ward (2016) stated, “Almost all have adopted new teacher evaluation systems” (p. 206). Countrywide teacher evaluation was now a standardized function.

### **TEACHNJ Act of 2012**

TEACHNJ was passed unanimously by the New Jersey State Legislature and was signed into law by Governor Christie on August 6, 2012 (Paxton, 2016). The legal (short) title of the law is C.18A:6-117, and the acronym stands for Teacher Effectiveness and Accountability for the Children of New Jersey. The bill starts with an opening declaration stating that the law was not going to provide new additional funding. It is written, “Existing resources from federal, State, and local sources should be used in ways consistent with this law” (C.18A:6-117, p. 1). Also stated in Article 1 of the legislation is that the goal of the bill 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” (C.18A:6-117, p. 1). In Article 2 it is revealed that the New Jersey Supreme Court found that a “multitude of factors play a vital role in the quality of a child’s education” (C.18A:6-117, p. 1). It goes on to mention that one of those factors is teacher effectiveness.

The law defines evaluation as:

a process based on the individual’s job description, professional standards and Statewide evaluation criteria that incorporates analysis of multiple measures of student progress and multiple data sources. Such evaluation shall include formal observations, as well as post conferences, conducted and prepared by an individual employed in the district in a supervisory role and capacity and possessing a school administrator certificate, principal certificate, or supervisor certificate. (C.18A:6-117, p. 1)

In three different sections the same exact definition of effective and highly effective is offered:

For purposes of this subsection “effective” or “highly effective” means the employee has received an annual summative evaluation rating of “effective” or “highly effective” based on the performance standards for his position established through the evaluation rubric adopted by the board of education and approved by the commissioner. (C.18A:6-117, pp. 5, 7, & 9).

And the definition for ineffective and partially effective is stated as:

For the purposes of sections 14 through 18 “ineffective” or “partially effective” means the employee receives an annual summative evaluation rating of “ineffective” or “partially effective” based on the performance standards for his position established

through the evaluation rubric adopted by the board of education and approved by the commissioner. (C.18A:6-117, p. 8)

TEACHNJ significantly altered the process for obtaining tenure as well the method for filing tenure charges (the process of dismissing of teachers that have obtained tenured status). Prior to the bill tenure charges could be obtained if a board of education could prove one of the following: (a) inefficiency, (b) incapacity, (c) conduct unbecoming a teaching staff member, or (d) some other just cause (§ 18A:6-10). Prior TEACHNJ many considered that the process of removing a tenured teacher was too burdensome and costly, as it involved a local board's superintendent, in conjunction with state commissioner of education, bringing charges before a state administrative law judge (Paxton, 2016).

TEACHNJ is 18 pages long and has 30 articles with the last two pages being a conclusion or summary "STATEMENT" (C.18A:6-117, p. 16). The word tenure, or the subject of teacher dismissal, can be found in exactly half of the articles: (a) three, (b) four, (c) five, (d) six, (e) seven, (f) 10, (g) 11, (h) 12, (i) 16, (j) 17, (k) 18, (l), (m) 23, (n) 24, (o) 28, and (p) 29 (a repeal of C.52:14B-10.1, the last bill concerning teacher tenure). Articles 6 and 7 are together two and a half pages long and detail the new provision of tenure in depth. On pages four and five (identical wording within separate text) the new tenure description is provided, "Three consecutive academic years, together with employment at the beginning of the next succeeding academic year with the first rating being received on or after the completion of the second year of employment" (C.18A:6-117). This added an additional year that faculty (and administrators) must be employed before tenure is received. In an analysis of the law, Paxton (2016) pointed out that "the Act removed tenure hearings from the OAL (Office of Administrative Law) and placed them in the hands of appointed arbitrators" (p. 401).

Articles 8 and 9 deal with issues of dismissing teachers when there is a reduction in force (e.g., fewer teaching positions needed for reasons such as a lowering of the student body).

Article 10 allows for teachers to retain tenure when taking positions in underperforming schools.

Article 13 stands alone just to identify the definition of ineffective and partially effective. Article

14 mandates and outlines the components of a school improvement panel. Article 19 deals with providing effective mentor teachers for first-year teachers, and in Article 20 it is stated:

A board of education shall provide additional professional development for any teaching staff member who fails or is struggling to meet the performance standards established by the board, as documented in the teaching staff member's annual summative evaluation. The additional professional development shall be designed to correct the needs identified in the annual summative evaluation. (C.18A:6-117, p. 12)

Article 21 mandates that school districts submit their evaluation model annually to the commissioner of education, and Article 22 stretches over a full page outlining the requirements of the evaluation models. This section of TEACHNJ starts by stating that the four categories of (a) ineffective, (b) partially effective, (c) effective, and (d) highly effective must be included in any evaluation model and that the models must use “multiple objective measures” (p. 12), “multiple measures of practice” (p. 13), and include “multiple observations” (p. 13). Article 25 states that charter schools must adhere to Articles 10, and 13–22. Articles 26 and 28 state that collective bargaining can only pertain to components that are not in conflict with the law. In Article 27 it is written that the Department of Education will provide the implementation funding, Article 30 sets the 2013-2014 school year as when the law will take effect.

The bill includes five articles that deal with the new teacher evaluation mandates, two concerning professional development, and 18 articles outlining the new changes to tenure and seniority provision. The remaining five articles are perfunctory in nature. The bill only provided for funding to effectuate its mass overhaul of state education law. Callahan and Sadeghi (2012-2013) explained, “The legislative intent is to make it more difficult for teachers to earn tenure and easier for school districts to eliminate underperforming teachers. The underlying assumption is that underperforming schools underperform because of bad teachers” (p. 1).

### **Standards Based Evaluation in the TEACHNJ Era**

In the post NCLB era public education policies involved measures to hold teachers accountable for student achievement; New Jersey was not exempt. The mainstreaming of accountability, resulting in a total revamping of New Jersey teacher evaluation, was addressed specifically in the TEACHNJ Act of 2012. This was the genesis of the four-tiered rubric for summative teacher evaluations. After 2 years of tinkering and piloting, 2013-2014 marked full implementation of standards-based teacher evaluation in New Jersey with 18 different teacher evaluation models for school districts to choose from. While each school district was free to choose which evaluation model they were to use, all of the models had to: specifically clarify teacher expectations, be used to inform and improve professional development, and provide meaningful feedback. All models had to also incorporate the four tiers of the new teacher rating scale (a) ineffective, (b) partially effective, (c) effective, or (d) highly effective, derived from combining a teacher’s observation scores conducted by his or her evaluating administrator with SGP and SGO scores, meeting the multiple data sources that are needed to be considered standards-based evaluation, according to Olivia and Pawlas (2004).

## **Administrator Time**

Pity the poor principal! He must be a manager, supervisor, psychologist, financial wizard, master of law, public relations specialist, public speaker, school and community leader, a first aid specialist; and through it all, he must be a good guy as well. He must be understanding, fair, reasonable, flexible, patient, stable, and always available – at school and elsewhere. (Moody, 1968, p. 543)

One of the first ever studies on principal behaviors was conducted by Feelhaver in 1927. The author studied high school principals in Nebraska and listed different components of principal duties (including the actual teaching of classes), and the only category related to teacher evaluation was “visiting” classrooms. However, Feelhaver’s research was aggregated by school size (total number of students), and it was found that as the number of students in a school increased, the number of classes the administrator taught went down. Student and teacher/administrator ratios had an influence on the tasks that principals undertook.

Without taking into account school size, as measured by the total number of attending students, Norton (1972) studied problems that principals faced in the daily operations of running their schools. The author made a hierarchical list of the 10 most common issues that administrators faced running their schools; and third on that list was “Problems of Teacher Personnel” (p. 456). Issues concerning teachers had become one of the more prominent problems that principals face, and the sheer number of tasks that administrators had to perform had grown considerably.

What is expected of administrators has never waned, and the call for educational accountability spurred on by *A Nation at Risk*, only added to the demands being placed on them. Federal and local policies forced administrators to become statisticians that had to analyze VAAs

and SGPs in the era of streamlined standardized teacher evaluation. NCLB forced principals to go through the process of ensuring that their teachers and schools met the “highly qualified” standard. With the AARA and RTTT grant program, administrators had to do the work of proving that student achievement was linked to teacher evaluation. The ultimate result of all the federal initiatives in New Jersey was the passing the TEACHNJ bill of 2012.

In a qualitative analysis of 30 principals, Donaldson (2011) pointed out that “states have increased their focus on the work of the principal in large part due to Race to the Top and other funding priorities initiated by the U.S. Department of Education and supported by Congress” (p. 1). Around the same time Grissom and Loeb (2011) studied managerial skills of principals using many factors. They started by administering an online survey to 314 school principals and 585 assistant principals (that ranked their principals) in the Miami-Dade County Public Schools. In this survey they included 42 “job tasks common to the principalship and asked to rate how effective they were at conducting each task in their current school” (p. 1095). The 42 tasks were placed into five categories: instructional management, internal relations, organization management, administration, and external relations. They ran factor analyses between various principal and school characteristics (gender, principal experience, principal highest degree, school type, and school size) and the 42 “time management” actions within the five categories. As additional variables, they included student achievement scores on the Florida Comprehensive Assessment Test and results from a parent survey. Of the various conclusions derived, most pertinent was that a majority of a principal’s time is *not* spent on teaching and learning activities.

Donaldson (2011) concluded, “Time” and “A limited opportunity to observe and document representative teaching” (p. 17) were two of just four factors that principals felt hindered their ability to effectively evaluate teachers. Grissom and Loeb (2011) stated that little

research exists examining time management and school administration; however, they did cite the components of general time management (outside the realm of educational administration) in the lead up to their study on principal time management skills. They used Britton and Tesser's (1991) Time Management Questionnaire—modifying it slightly to measure principals' time management skills. The survey included three types of time management, “short-range planning, long-range planning, and time attitudes” (p. 780) and 21 total descriptors of time management skills. In their conclusion, the authors found that time management was important in many different ways: better time management equaled being able to spend more time on instructional leadership (which they found to be what principals considered most important); there was a relationship between better time management and lower job stress; and, time management is related to productive work behaviors and positive self-assessment.

### **Instructional Leadership**

In a study of the principalship from the 1920s through the 1990s, Beck and Murphy (1993) identified the 1980s as the genesis of the practice of instructional leadership. Hallinger (2011), in a review of 130 doctoral dissertations (EdD and PhD), stated that instructional leadership was widely accepted by policy makers on national and international levels. The dissertations that Hallinger (2011) included all used the Principal Instructional Management Rating Scale (PIMRS), which was the “most widely used instrument in school leadership research over the past 30 years” (p. 273). The framework of PIMRS has 10 components in three categories: Defining the School Mission, Managing the Instructional Program, and Developing the School Learning Climate. Hallinger (2011) noted that the scholars in his study and previous research chose to study administrators at the elementary level.

In the review, Hallinger (2011) identified nine different models that were used to measure

PIMRS, for example, principal demographics and instructional leadership, direct effects of instructional leadership on school variables, and reciprocal effects of instructional leadership on school variables and/or outcomes. Grissom and Loeb (2011) cited Hallinger and Murphy in referring to instructional leadership as “‘anything and everything’ (p. 217) principals might do to support classroom learning” (p. 1093). Grissom and Loeb (2011) also stated that the PIMRS model was used in over 100 studies (including dissertations). Marzano, Waters, and McNulty (2005) found that instructional leadership was one of the most frequently researched educational leadership paradigms, even though they asserted that it was not always well defined. Hanushek (2005) included the notion that instructional leaders had to make decisions within the context of economic and political policies. The literature on instructional leadership is dense with a long list of tasks that principals must take on to meet *any* definition of effective instructional leadership

### **Transformational Leadership**

Hallinger (2003) noted that transformational leadership and instructional leadership made up the two most measured leadership models. The post *A Nation at Risk* era brought top-down approaches to school leadership, which encompassed instructional leadership. Transformational leadership was a reaction that followed (Hallinger & Heck, 1996). As a transformational leader, a principal qualitatively elicits and examines his or school and induces change in a bottom-up fashion. Empowerment is the key word when it comes to transformational leadership (Hoy & Miskel, 2008; Razik & Swanson, 2001). Van Knippenberg and Sitkin (2013) criticized transformational leadership stating that there is no clear way to define exactly how to apply the concept in any uniform way.

Transformational leadership is not based in accomplishing a measurable checklist; it is based on building relationships and facilitating a climate where school needs are met with a

collective vision. Bass and Riggo (2006) provided some of influences that successful transformational leaders achieve. They provide an atmosphere where stakeholders: (a) work together to create a shared vision; (b) facilitate all followers to be situational problem solvers; and (c) demonstrate a model of trust, integrity, and dedication. In bringing school personnel together, transformational leaders coach, mentor, and support all the members of their organization: all while creating a trusting environment (Bass & Riggo, 2006). Transformational leadership involves a principal being a deeply dedicated leader that takes on whatever role necessary in empowering followers to collectively and individually follow a shared vision for success. Research suggests that it takes time and dedication in order to be successful.

### **Theoretical Framework**

#### **Production Function**

Production function is an economic theory that “explains a basic technological relationship between scarce resources, or inputs and output (Gordon & Vaughn, 2011, p. 25). There are various ways to set up inputs in production function experiments, but essentially one can provide fixed inputs, variable inputs, or a combination of the two. Outputs can be measured in the short run (maturity fulfillment in a year’s time, or one theoretical time period, e.g., a school year) or the opposite, long run, where no inputs are fixed (Gordon & Vaughn, 2011).

The genesis of production function theory dates back to 1767 when Anne Robert Jacques Turgot postulated the “concept of diminishing returns” (Gordon & Vaughn, 2011, p. 26). Technological advancements in the world changed the inputs and outputs that could be measured, ergo, the uses of production function evolved and expanded. The Cobb-Douglas (CD) method of production function, first published in a 1928 paper by Charles Cobb and Paul Douglas (Berndt & Christensen, 1973), and later noted as the most widely used form of

production theory by Ringstad in 1967, outlined a specific framework for the theory. The CD production function method provided for an *exponential* value to be derived of each input variable, which are then totaled up to a maximum input from zero to one; after which they are computed to measure an output (Jia, Long, Wang, Yan, & Kang, 2016). This allowed for the input variables to substitute each other over the time period of the production function measurement.

Robert Solow's constant elasticity of substitution (CES) came about in 1956, built upon CD but was distinguishably different. As it is titled, the inputs do not just substitute each other, but the substitutions infinitely vary before becoming a constant value used in measurement (Gordon & Vaughn, 2011; Mansanjala & Papageorgiou, 2004; Solow, 1956). In evaluating the merits of CD and CES, Klump, McAdam, and Willman, (2012) stated that because CES allows for a normalization of *elastic* input variables (instead of only a unitary value in CD models), and that "income shares (output) change over time" (p. 792), CES production function models were the superior of the two.

### **Education Production Function**

The use of production function theory was limited to the economic realm until it was applied to measure educational productivity (with standardized test scores as the "output") in the Coleman Report of 1966, titled *Equality of Educational Opportunity* (Coleman et al., 1966). The Coleman Report mandated by the Civil Rights Act of 1964, and published in 1966, was a 749-page grand-scale sociological experiment that, at its conclusion, found socioeconomic status best predicts educational outcomes; therefore, minority students were disadvantaged when compared to their White counterparts. Facility, curriculum, and faculty were all factors that had little, if any, statistical relationship with student outcomes. Poverty had to be at the heart of any

educational reform. Mayeske et al. (1972) conducted a second large-scale government funded study titled, *A Study of Our Nation's Schools*. It was found that “on the whole, the influence of school cannot be separated from that of the student’s social background--and vice versa” (p. xiv).

A reaction to Coleman et al. (1966) and Mayeske et al. (1972) was a backlash of studies attempting to find data to the contrary. Regardless of the validity of the post-Coleman studies, production function theory was now the tool of tools to measure education outcomes, with standardized test scores (often labeled “achievement”) as the accepted output. In a publication titled *On Equality of Education Opportunity*, edited by the late Senator Daniel Patrick Moynihan and the late eminent educational statistician Frederick Mosteller, the authors Hanushek and Kain (1972) used the tenets of production function theory to disrupt the findings of Coleman Report, contesting the methodology point by point and citing the Hanushek’s dissertation titled, *The Education of Negroes and Whites*, in the process.

Bowles (1970), writing for the National Bureau of Economic Research (a private non-profit organization), and prior to the large-scale Coleman Report rebuttals, was alarmed that education spending was “increasing at a rate more than twice that of the economy” (p. 11). To address the economic concern, and to help define education production function (EPF), Bowles evaluated the data in the Coleman Report using standardized test achievement as a proxy for economic indicators outside of the school. Bowles used various variables under the categories of (a) non-school, (b) school environment, (c) teacher quality, (d) teacher quantity, (e) school facilities, and (f) student attitudes. One of the only results the author reported is controversial; genetics play (using IQ) a role in the economic output. Contrary to the Coleman Report, Bowles reported, “We are a long way from estimating a satisfactory educational production function” (p. 51).

Hanushek defined EPF as “a function that relates various inputs to education including those of families, peers, and schools to the maximum level of student achievement that can be obtained” (2010, p. 407). In a 1981 analysis of 130 separate analyses, across 29 published studies, Hanushek claimed that no relationship between school expenditures and achievement *could* be found to exist, and, therefore, “Throwing Money at Schools” (title of article; p. 19) was futile.

Hanushek (2010) stated that EPFs are “typically some form of regression analysis” (pp. 45–46). This time the author included 187 experiments found in 38 journals or books (p. 46). Hanushek concluded that family background diminishes any relationship between school expenditures and academic achievement thusly claiming that any increase in education spending will not result in achievement gains in the future.

Using EPF to confirm the Coleman Report findings, Greenwald, Hedges, and Laine (1996) denied the two Hanushek studies above plus two others (conducted in 1986 and 1991). They stated that Hanushek had used “vote counting” (p. 362), which was a method that was not used when high level results are expected. They used a meta-analysis of 60 research studies that either controlled for socioeconomic status or were longitudinal in nature. Succinctly affirming their method, Greenwald et al. concluded that there is a significant relationship between school resources (expenditures) and academic achievement.

Using the *National Educational Longitudinal Study of 1988* data on 10th-grade mathematics scores as their source, and noting that teacher salaries and benefits took the lion’s share of educational budgets, Goldhaber and Brewer (1997) used EPF to measure whether unobservable factors impact analyses. While admitting unobservable factors are important, the authors claimed that omitting them do not bias results of EPF experiments using only observable

factors.

Although the overwhelming amount of EPF studies use student achievement on standardized tests as the output, on a smaller scale EPF has been used with *grades* (college economic course) as the output. Borg and Shapiro (1996) inputted Myers-Briggs personality types (for both professor and student), ethnicity (actually only African American/not African American), teacher teaching style, high school GPA, SAT scores, and previous college credits to determine the output of the grade achieved in a college economics course.

Douglas and Sulock (1995) conducted a similar study with no professor traits being used but with other inputs such as gender, age, and if night classes were being taken by the student. Douglas and Sulock (1995) cited Heckman's (1991) two-step method to correct for sample-selection bias, in order to control for their students that dropped the course. Heckman explained that "simple regression techniques to estimate behavioral functions free of selection bias (can be used) in the case of a censored sample" (p. 162).

Studying factors that influence teacher summative evaluation ratings scored by their evaluating administrators is a phenomenon that has not been researched before. Provided in the New Jersey School Performance Report, data on various student and school variables can be found; in particular for this study the administrator/student ratio for every public school in New Jersey are available. The New Jersey Staff Evaluation spreadsheet provides school level results for all teachers using the (a) ineffective, (b) partially effective, (c) effective, or (d) highly effective rating scale. A chronological study on teacher evaluation was conducted that demonstrated the current countrywide use of standardized teacher evaluation systems.

In New Jersey standardized teacher evaluation came to be with the TEACH NJ Act of 2012. Administrator Time, Instructional Leadership and Transformational Leadership were

studied as key factors that influence how an administrator goes through the process of teacher evaluation. Lastly, production function theory, and more specifically, education production function theory was examined thoroughly as input/output analyses were used in this study. Administrator/student ratios (while controlling for student and school variables) were inputted, and the output was the school-level percentages of teacher rating results as scored by their evaluating administrators.

Chapter III of this study outlines the analyses conducted in which the output (teacher summative evaluation results scored by their evaluating administrators) was determined by the input (administrator/student ratios while controlling for student and school variables). As cited by Sammarone (2014), Aigner and Chu (1968) asserted, “For the goal of fitting a function through a series of observations on firms for outputs and several inputs, this implies that an ‘average’ function is obtained” (p. 67).

## CHAPTER III

### METHODOLOGY

I conducted this quantitative research study to explain the influence that administrator/student ratios had on the summative teacher practice scores that teachers receive based upon classroom observations by school administrators. In addition, I aimed to explain the amount of variance in summative teacher ratings accounted for by administrator/student ratios when controlling for other student- and school-level factors. There is an absence of research that measures the influence that school and student factors have teacher summative evaluation scores (and more specifically the administrator/student ratios).

#### **Research Design**

I used a correlational, cross-sectional, explanatory design with quantitative methods. Using administrator/student ratios and a combination of school and student variables, I attempted to find “causal factors that produce(d) a change” (Johnson, 2001, p. 9) in teacher summative evaluation ratings scored by their evaluating administrators. First, a correlational design was used to explain the relationship that exists between administrator/student ratios and teacher summative evaluation results scored by their specific evaluating administrators. Correlational analysis is a technique used to “describe the association between two or more quantitative variables” (Schutt, 2015, p. 345). The quantitative methods served to determine the strength and direction of any relationships that existed between the independent variables and teacher evaluation results as reported by the evaluating administrators. “Statistics based on regression and correlation are used frequently in social science and have many advantages over cross-tabulation” (Schutt, 2015, p. 345). While going beyond a solely descriptive analysis, it would not have been appropriate to run an experiment (or series of experiments) looking for a cause and

effect relationship between just one independent variable and one dependent variable. Turnamian (2012) wrote:

Within the field of social sciences, most research problems are not easily examined through experimentation. Therefore, correlational studies are one of the frequently used research designs in the social sciences and can limit research from finding causality between two variables. Non-experimental causal comparative research designs do attempt to provide evidence of cause and effect relationships between variables and can be seen as a non-experimental research design which may identify causality (p. 100).

In this study, after running a correlation matrix, a simultaneous multiple regression model was used in order to specify which variables influenced teacher evaluation scores. Then I was able to then organize the variables by the significance and strength of the correlation. The strongest variables were used to run separate regression models for each grade level.

Then hierarchical regression models, derived from the correlational design and simultaneous multiple regressions, were used to specifically examine the influence of the independent variable on the dependent variable. Hierarchical multiple regression models allow various independent variables to determine the dependent variable (Leech, Barret, & Morgan, 2015). The goal was to discern which student variables (percentage of ELL students, percentage of students qualifying for free or reduced lunch, and percentage of students with an IEP) and school variables (administrator/student ratio and per-pupil expenditure) had a statistically significant relationship to teachers being scored (a) ineffective/partially effective, (b) effective, or (c) highly effective on summative evaluations scored by their specific evaluating administrators.

## **Research Questions**

**Research Question 1:** What is the influence of school administrator/student ratios at elementary schools containing grades PreK–6 (as identified by containing three consecutive grades or more, with no class above the sixth grade) on the percentage of teachers being rated effective/highly effective on their summative evaluations scored by the administrators at those schools?

**Research Question 2:** What is the influence of school administrator/student ratios at elementary schools containing grades PreK–6 (as identified by containing three consecutive grades or more, with no school class above the sixth grade) on percentage of teachers being rated highly effective on their summative evaluations scored by the administrators at those schools?

## **Null Hypotheses**

**Null Hypothesis 1:** No statistically significant relationship exists between school administrator/student ratio and the percentage of teachers scored effective by administrators at those schools.

**Null Hypothesis 2:** No statistically significant relationship exists between school administrator/student ratios and the percentage of teachers scored highly effective by administrators at those schools.

## **Sample Population/Data Source**

The sample used included public elementary schools within the 21 counties of New Jersey. Excluded were charter and special education schools. For a school to be included the following criteria had to be met:

A: The schools were classified as public.

B: The schools were PreK–6 schools that contain three consecutive grades or more, with no class above the sixth.

C: The schools reported (and were published data on the NJDOE website) the number of teachers that received effective AND highly effective scores on their summative evaluations.

The number of schools that met the criteria was ( $n = 242$ ).

### **Independent Variables**

#### **Administrator/Student Ratio**

The raw total of students in the administrator/student ratio of each school was used and inputted as a numeric variable.

#### **District Per-Pupil Expenditure**

Each district's per-pupil expenditure, in dollar amounts, was rounded to the nearest thousand and inputted as a numeric value.

#### **Percentage of Economically Disadvantaged Students**

The percentages of economically disadvantaged students were entered as a whole number and inputted as a numeric value.

#### **Percentage of Limited English Proficient Students**

The percentages of limited English proficient students were entered as a whole number and inputted as a numeric value.

#### **Percentage of Students with a Disability**

The percentages of students with a disability were entered as a whole number and inputted as a numeric value.

## **Dependent Variables**

The percentages of teachers scored by their building level administrator as: (a) ineffective/partially effective, (b) effective, and (c) highly effective were the dependent variables. These score categories were taken from the rubric outlined in the TEACHNJ Act of 2012, which was created to be the standardized statewide teacher evaluation tool (Ch. 26; C.18A:6-117). A descriptive analysis of the data demonstrates that across the board very few teachers are scored in the first category (ineffective/partially effective); however, there was a noticeable variance between the effective and highly effective categories. The evaluation score data were demonstrated in the percent for each category: (a) ineffective/partially effective, (b) effective, and (c) highly effective

## **Data Collection**

The data for this study were retrieved from three New Jersey Department of Education websites: (<http://www.state.nj.us/education/data/staff/>, <https://rc.doe.state.nj.us/ReportsDatabase.aspx>, and <http://www.nj.gov/education/guide/2016/>). The three Excel spreadsheets (Staff Evaluations 2015-2016, 2015-2016 School Performance Reports, and The Taxpayers Guide to Education Spending 2016, respectively) were downloaded and saved to data files. A unique identification code was made by merging the county district and school numeric identifiers. Data from public PreK–6 schools that contain three consecutive grades or more, with no class above the sixth with reported teacher evaluation scores in the effective and highly effective categories were used in this study (see Table 2). Schools that did NOT have reported scores for the number of teachers in both the effective and highly effective categories were removed. Then various sorting options were applied to remove all schools that

did not meet the grade-level requirements or were considered charter or alternative. The schools that met criteria were then arranged in alphabetical order by county, district, and school.

Table 2

*Data Retrieved from the NJDOE and Utilized in the Excel Spreadsheet*

County Name
District Name
School Name
School Grade Span
Administrator/Student Ratio
District Per-Pupil Expenditure
Percentage of Students Economically Disadvantaged
Percentage of Students on Limited English Proficient
Percentage of Students with a Disability

The tallies for summative evaluation scores that teachers received at the 242 schools were added to the spreadsheet. Ineffective and partially effective counted as one category together because of the low frequency of those two ratings, whereas effective and highly effective were stand-alone categories. The clean and formatted data were then imported into Statistical Package for the Social Sciences (SPSS, Version, 19) statistical software.

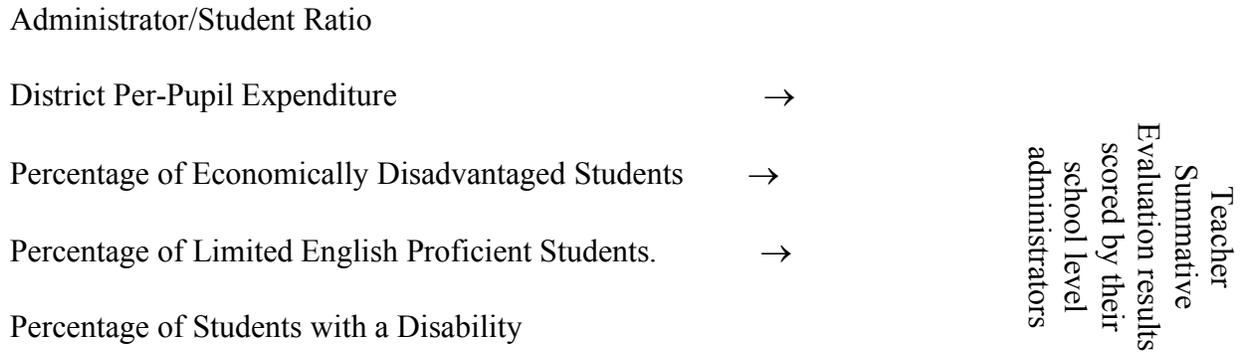


Figure 1. Simultaneous regression framework.

### Data Analysis

First, I explored descriptive statistics to identify means, ranges, and standard deviations of the variables. I also checked the skewness of the dependent variables to ensure they met the assumption of normality as required when using simultaneous regression.

Next, I ran a Pearson correlation test and created a correlation matrix. The purpose of the correlation matrix was to investigate the strength and direction of any initial relationships that existed. The correlation matrix was also helpful in identifying potential causes of multicollinearity.

After running a correlation matrix, a simultaneous multiple regression was used in order to specify which variables influenced teacher evaluation scores. There were five predictor variables and one dependent variable in the two analyses. Using a statistical significance threshold of  $p \leq .05$  and seeking a medium effect size, I needed to meet an appropriate minimum sample size. Miles and Shelvin (2010) graphed predictor variables, and the corresponding required sample size for regressions; according to their assertion, I would need a sample of at least 100 when having six predictor variables. Field (2009) suggested using the formula of  $104 +$

$k$ , with  $k$  being the number of predictor variables. I had a sample of 242, well above Miles and Shelvin's (2010) required sample size of 100 and 133 datum points above the sample size of 109 suggested by filling in five (my  $k$ ) into Field's (2009) formula. My sample would have provided the means needed to result in a statistically significant effect size (.50) at the 95% confidence interval.

All five predictor variables were used to create a correlation coefficient matrix with the teacher rating scores (dependent variable); then, I ran a simultaneous multiple regression. Pallant (2010) stated that simultaneous multiple regressions can be powerful in comparison to an individual regression, as more than one predictor variable are used to determine the relationship with the dependent variable. Thus, the researcher is able to determine which variable best predicts the dependent variable. In doing this I sought to find any statistically significant variables. In order to verify my results, I ran a backwards simultaneous regression analysis to verify the results from the first model. The dependent variable total proficient and advanced proficient were run against all of the independent variables (See Model as Table 3).

Table 3

*Simultaneous Regression Model*

Teacher summative evaluations scored by school administrators	All school and student variables	Administrator/Student ratio
		District Per-Pupil Expenditure
		Percentage of Economically Disadvantaged Students
		Percentage of Students Limited English Proficient
		Percentage of Students with a Disability

I stopped after running the simultaneous regressions because only one of the five variables was found to be a statistically significant predictor for both dependent variables (EFFECT% had two significant variables; however, the *p* value of the second variable, PerPupil\$, was .049) of the evaluation scores; the *R* squared also revealed that the strength of the variable was very weak.

**Instrumentation**

As of May 1, 2017, education officials at the New Jersey Department of Education approved 26 evaluation instruments that were used throughout the state (with 13 additional districts choosing not to have their model published):

- 5D+TM Teacher Evaluation Rubric
- Charlotte Danielson: The Framework for Teaching (2007 Edition)
- Charlotte Danielson: The Framework for Teaching (2011 Edition)
- Charlotte Danielson: The Framework for Teaching (2013 Edition)

Charlotte Danielson: The Framework for Teaching Instructionally Focused Edition (2013)  
Classroom Assessment Scoring System (CLASS)  
Classroom Strategies Scale Model  
Focal Point Teaching Practice Model  
IMPACT: The DCPS Effectiveness Assessment System For School-Based Personnel  
H.E.A.T./Danielson Teacher Evaluation Instrument  
Insight Core Framework  
Kenilworth Teacher Evaluation Instrument  
Lenape Regional Teacher Evaluation Instrument  
Marzano’s Causal Teacher Evaluation Model  
Mid-Continent Research for Education and Learning (McREL) Teacher Evaluation Standards  
North Star Academy Teacher Evaluation Rubric  
Pearson Framework for the Observation of Effective Teaching  
Rhode Island Model: Teacher Evaluation & Support System (Edition II)  
Stronge Teacher and Leader Effectiveness Performance System  
Teacher Evaluation and Improvement Instrument  
The College-Ready Promise Teaching Framework(CRPTF)  
The Marshall Rubrics  
The Newark Public Schools  
The New Jersey LoTi Teacher Evaluation  
The SmartStart TeachElite Evaluation System  
The Thoughtful Classroom Teacher Effectiveness Framework (NJDOE, n.d.g., pp.1–3)

On the Frequently Asked Questions portion of the NJDOE’s AchieveNJ Home page it is written:

Q: What are the requirements for selecting evaluation instruments?

A: Teacher (and principal) practice evaluation instruments must be approved by the New Jersey Department of Education. Current State-Approved Teacher and Principal Practice Evaluation lists have been posted. We recognize that districts may wish to change elected instruments in the future as new and updated instruments become available. New RFQ processes will be communicated directly to districts. Districts share information about instrument changes through annual evaluation reporting procedures. (NJDOE, n.d.g).

In the text of the Teach NJ Act of 2012 it is written (Ch. 26; C.18A:6-117):

The standards at a minimum shall include: (1) four defined annual rating categories for teachers, principals, assistant principals, and vice-principals: ineffective, partially effective, effective, and highly effective; (2) a provision requiring that the rubric be partially based on multiple objective measures of student learning that use student growth from one year's measure to the next year's measure; (3) a provision that allows the district, in grades in which a standardized test is not required, to determine the methods for measuring student growth; (4) a provision that multiple measures of practice and student learning be used in conjunction with professional standards of practice using a comprehensive evaluation process in rating effectiveness with specific measures and implementation processes. Standardized assessments shall be used as a measure of student progress but shall not be the predominant factor in the overall evaluation of a teacher; (5) a provision that the rubric be based on the professional standards for that employee; (6) a provision ensuring that performance measures used in the rubric are linked to student achievement; (7) a requirement that the employee receive multiple observations during the school year which shall be used in evaluating the employee; (8) a provision that requires that at each observation of a teacher, either the principal, his designee who shall be an individual employed in the district in a supervisory role and capacity and who possesses a school administrator certificate, principal certificate, or supervisor certificate, the vice-principal, or the assistant principal shall be present; (9) an opportunity for the employee to improve his effectiveness from evaluation feedback; (10) guidelines for school districts regarding training and the demonstration of competence on the evaluation system to support its implementation; (11) a process for ongoing monitoring and calibration of the observations to ensure that the observation protocols are being implemented correctly and consistently; (12) a performance framework, associated evaluation tools, and observation protocols, including training and observer calibration resources; (13) a process for a school district to obtain the approval of the commissioner to utilize other evaluation tools; and (14) a process for ensuring that the results of the evaluation help to inform instructional development. (p. 9)

All 26 evaluation rubrics approved by the NJDOE, and deemed valid by the state

(including those used in this study), had to meet the 14 criteria outlined in the law.

The results presented in Chapter IV will start with an analysis of the descriptive data, including skewness kurtosis testing. Then a correlation matrix will be run. After which, an “enter method” simultaneous multiple regression, including all five independent variables run at once, will be conducted. This will be done to glean significant predictor variables. When it is demonstrated that statistical significance was not found between the predictor variables and the dependent variables, then a response to the null hypotheses will be made. The analyses were used to find:

1. the overall statistical significance in the ANOVA table and
2. the  $R$  squared and  $R$  squared changes to determine which variables contribute most to the  $R$  squared value.

## CHAPTER IV ANALYSIS OF THE DATA

### **Introduction**

The purpose for this correlational, cross-sectional study was to explain the influence that administrator/student ratios had on the summative evaluation ratings that teachers received from their evaluating administrators. The data analyzed included administrator/student ratios with controlled student and school variables; the sample was of 242 K–6 elementary schools that had reported teacher evaluation percentages in the (a) effective and (b) highly effective categories. I aimed to provide research-based evidence on the influence of the TEACHNJ Act of 2012 and the staff evaluation provisions found within by the New Jersey State Legislature. The results from this study serve to demonstrate that the school and student factors studied do not influence the summative evaluations scores the teachers receive from their evaluating administrators.

### **Independent Variables**

Existing research suggested variables that influence the everyday actions of administrators that evaluate their teaching staff. Grissom and Loeb (2011) demonstrated that administrators pointed to (a) overall time and (b) time to conduct observations as two of four important factors important to their success. This makes administrator/student ratios an important variable as the amount of students in a school impacts a principal's available time. In 1991, the Individual with Disabilities Education Act (IDEA) mandated that students with disabilities participate in all standardized tests—district or state level (Yell, Katsiyannis, Collins, & Losinski, 2012), and in 2001 NCLB did the same for LEP populations (Wright & Li, 2008). Ever since The Coleman Report in 1966, SES has been identified as having a universal influence on every facet of public school.

Table 4

*Variables and Names of Independent Variables*

Variable	Label	Description
Administrator/Student ratio	A/SRatio	Administrator/Student ratio
Percentage of economically disadvantaged students	Poverty%	Percentage of students eligible for free or reduced lunch
Percentage of students with a disability	Disability%	Percentage of students with a disability
Percentage of limited English proficient students	LEP%	Percentage of limited English proficient students
District per-pupil expenditure	PerPupil\$	The amount of money spent per pupil rounded to the nearest thousand.

**Dependent Variables**

The New Jersey Department of Education requires every school district to report specific school and student data, as well as teacher summative evaluation scores. The school and student data and staff evaluation results are found on the NJ School Performance Report and the NJ Staff Evaluation Spread Sheet respectively. The dependent variables for this study were the percentage of teachers that received (a) effective and (b) highly effective summative teacher scores by their school administrators. All data regarding schools other than K–6 elementary schools were deleted as well as data pertaining to schools that did not report both the effective and highly

effective totals for teachers at those schools (if the  $n$ -size is less than 10 then the scores for that category must be suppressed). I labeled the percentage of teachers receiving effective scores as EFFECT% and the percentage of teachers receiving highly effective scores as HIGH%.

### **Procedure**

I used quantitative methods to explain the influence of administrator/student ratios, and other school and staff factors, on the percentage of teachers being rated effective and highly effective on their summative evaluations scored by the administrators at those schools. First, I explored descriptive statistics to identify means, ranges, and standard deviations of the included variables. I also checked the skewness of the independent variables to ensure they met the assumption of normality, as required when using simultaneous regression.

A positive skewness can suggest that the data are tailed (spread out) to the right, and the mean falls to the right of the median and conversely for a negative skewness score. Skewness may indicate that outlying data are impacting the distribution. Field (2000) suggests that data with a skewness of an absolute value of 1 is acceptable, with normal distribution being reached the closer the value is to zero.

Next, I ran a Pearson correlation test and created a correlation matrix. The purpose of the correlation matrix was to investigate the strength and direction of any initial relationships that existed among the variables. The correlation matrix was also helpful in identifying potential causes of multicollinearity. Multicollinearity is present when independent variables are closely associated with each other, and their relationship could be the cause for an influence on the dependent variable.

After running a correlation matrix, a simultaneous multiple regression was used to explain which variables, if any, influenced summative teacher evaluation scores. There were five

predictor variables and one dependent variable used in the analyses. Using a statistical significance threshold of  $p \leq .05$ , I examined the model summary and ANOVA table of each model to look for statistical significance and an  $R$  squared value.  $R$  squared tells the researcher what percentage of the variation in the dependent variable is explained by the data in the regression model; an  $R$  squared of 1 equals a perfect and total explanation.

The ANOVA models demonstrated statistical significance. This prompted me to create and examine a coefficient matrix using all five predictor variables. Using a  $p$  value of  $< .05$ , I found that Poverty% and PerPupil\$ (at the  $p < .10$  for highly effective) were the only two variables that demonstrated statistical significance. Having a total variance of 6.3% and 6.2% for my two dependent variables, the data revealed that there was no statistical reason to move on to hierarchical regressions as demonstrated in the results below.

## **Results**

I calculated the means and standard deviations for the dependent and independent variables used in correlation and regression analyses (Table 5). The mean percentage of teachers that were rated as highly effective was approximately 49% with a standard deviation of 14.24. The percentage of teachers rated effective was approximately 51% with a standard deviation 14. The average percentage of students who qualified for free or reduced lunch was approximately 29% with a standard deviation of 24.82; the percentage of LEP students was approximately 6% with a standard deviation of 7.24; and the percentage of students with a disability was approximately 18% with a standard deviation of 5.82. The average per-pupil expenditure was approximately \$19,000, and the average student/administrator ratio was approximately 333 to 1.

Table 5

*Descriptive Statistics Table*

	<i>N</i>	<i>M</i>	<i>SD</i>
HIGH%	242	48.73	14.24
EFFECT%	242	50.82	14.28
A/SRatio	242	332.62	123.87
Disability%	242	17.99	5.82
Poverty%	242	28.81	24.82
LEP%	242	5.81	7.24
PerPupil\$	242	19157.025	2182.57

After that I calculated the descriptive statistics for the dependent variables. First, I computed the data for teachers being scored effective on their summative evaluations performed by their building administrators (Table 6). Using approximate values, the mean percentage of teachers that were scored effective was 51%, median 50%, and the standard deviation was 14%.

With a maximum of 83% and a minimum of 18% and a standard deviation of 14.27, the data revealed that almost all schools reported the percentage of teachers scored effective within the range of 21.5 % to 78.5 %. Approximately two thirds of the schools reported 36% to 64% of its teachers as being scored as effective. The data revealed that the percentage of teachers reported to be scored above 80% and below 20% was statistically not significant and negligible in practical terms.

Table 6

*Descriptive Statistics for Teachers Receiving Effective Scores on Summative Evaluations*

		Statistic	Std.Error
EFFECT%	Mean	50.81	.91
	95% confidence	Lower bound	49.01
	Interval for mean	Upper bound	52.62
	5% Trimmed mean		50.74
	Median		50.00
	Variance		203.80
	Std. deviation		14.27
	Minimum		19.00
	Maximum		83.00
	Range		64.00
	Interquartile range		22.25
	Skewness	.057	.156
	Kurtosis	-.734	.312

Next, I calculated the skewness. The skewness of the results was .057 demonstrating a normal distribution, as it is less than the absolute value of 1 and hovering around zero. The Shapiro–Wilk test (Table 7) revealed a significance value of .022, less than .05, suggesting that the distribution may not be normal, so I examined the skewness and kurtosis further to calculate *z*-scores. I did not need to find the *z*-score for the skewness being that the standard deviation was larger than the skewness (.156 and .057 respectively) and could only have a *z*-score even closer

to zero when using the standard deviation as the divisor. The kurtosis  $z$ -score was 2.353 ( $|-.734| / |.312| = 2.353$ ) and less than an absolute value of 3.29, which falls in the range of normality, according to Kim (2013).

Table 7

*Effective Evaluation Scores Normality Table*

	<b>Shapiro–Wilk</b>		
	Statistic	df	$p$
EFFECT%	.987	242	.022

Second, I computed the data for teachers rated as highly effective on their summative evaluations performed by their building administrators (Table 8). Using approximate values, the mean percentage of teachers that were scored effective was 49%, median 48%, and the standard deviation was 14%. The skewness was -.03 demonstrating a normal distribution, as it is less than the absolute value of 1 and hovering around zero. The kurtosis statistic was -.740.

With a maximum of 81% and a minimum of 17% and a standard deviation of 14.24 % the data revealed that almost all schools reported the percentage of teachers scored highly effective within the range of 20.5% to 77.5%. Approximately two thirds of the schools reported 35% to 63% of teachers as being scored as effective. The data also revealed that the percentage of teachers reported to be scored above 80% and below 20% was for practical purposes not significant and negligible.

Table 8

*Descriptive Statistics for Teachers Receiving Highly Effective Scores on Summative Evaluations*

		Statistic	Std. Error
HIGH%	Mean	48.73	.91
	95% confidence interval for mean	Lower bound Upper bound	46.92 50.53
	5% Trimmed mean	48.78	
	Median	48.00	
	Variance	202.79	
	Std. deviation	14.24	
	Minimum	17.00	
	Maximum	81.00	
	Range	64.00	
	Interquartile range	22.25	
	Skewness	-.030	.156
	Kurtosis	-.740	.312

The Shapiro–Wilk test (Table 9) revealed a significance value of .027, less than .05, suggesting that the distribution may not be normal, so I examined the skewness and kurtosis further to calculate  $z$ -scores. I did not need to find the  $z$ -score for the skewness being that the standard deviation was larger than the skewness (.156 and .03 respectively) and could only have a  $z$ -score closer to zero when using the standard deviation as the divisor. The kurtosis  $z$ -score was 2.372 ( $|-.740| / |.312| = 2.372$ ), and less than an absolute value of 3.29, which falls in the

range of normality, according to Kim (2013).

Table 9

*Effective Evaluation Scores Normality Table*

	<b>Shapiro–Wilk</b>		
	Statistic	df	<i>p</i>
HIGH%	.987	242	.027

### **Correlations**

Table 10 is a correlation matrix that includes all five independent variables with each other. There was a significant ( $p < .01$ ) relationship between Poverty% and LEP% of medium strength (.528) demonstrating, while not a high-level association, a further examination for multicollinearity could be conducted.

Table 10

*Pearson Correlations of the School and Student Independent Variables*

	A/S Ratio	Disability %	Poverty %	LEP %	PerPupil\$
A/S Ratio	1				
Disability %	.031	1			
Poverty %	-.058	-.173**	1		
LEP %	-.065	.221**	.528**	1	
PerPupil\$	-.077	.054	.011	-.008	1

*Note.* \*  $p < .05$ , 2-tailed. \*\* $p < .01$ , 2-tailed.

Table 11 is a correlation table measuring all five independent variables with the two dependent variables. A statistically significant relationship ( $p < .01$  threshold) was found between the percentage of students receiving free or reduced lunch and both dependent variables. The percentage of LEP students was also found to have a statistically significant relationship with both independent variables at the significance  $< .05$  threshold. These data, and lack of statistically significant relationships, foreshadowed what I was going to find in the simultaneous multiple regressions that followed.

Table 11

*Pearson Correlations of the School and Student Variables and Effective and Highly Effective Percentages*

	EFFECT%	HIGH%
A/SRatio	-.010	.005
Diasabilty%	-.097	.100
Poverty%	.208**	-.209**
LEP%	.136*	-.133*
PerPupil\$	-.126	.118

*Note.* \*  $p < .05$ . \*\*  $p < .01$ .

### **Simultaneous Multiple Regressions**

I ran the first model using all five predictor variables with the percentage of teachers that were scored effective as the dependent variable. The model summary is shown below (Table 12) with the  $R$  squared revealed as .063.

Table 12

*Percentage of Teachers Scored Effective Model Summary*

Model	<i>R</i>	<i>R</i> <sup>2</sup>	Adjusted <i>R</i> <sup>2</sup>	Std. error of the estimate
1	.251 <sup>a</sup>	.063	.043	13.96452

Predictors: (Constant), PerPupil\$, LEP%, A/SRatio, Disability%, Poverty%<sup>a</sup>

The ANOVA data (Table 13) show that the model was statistically significant with  $F = 3.182$ ,  $df = 5$ , and  $p = .008 = < .05$ .

Table 13

*Percentage of Teachers Scored Effective ANOVA Table*

Model		Sum of squares	<i>df</i>	Mean square	<i>F</i>	<i>p</i>
1	Regression	3102.186	5	620.437	3.182	.008 <sup>b</sup>
	Residual	46021.814	236	195.008		
	Total	49124.000	241			

*Note.* Dependent Variable: EFFECT%<sup>a</sup>

Predictors: (Constant), PerPupil\$, LEP%, A/SRatio, Disability%, Poverty%<sup>b</sup>

The coefficients table demonstrated the beta values of the independent variables. Beta values are used to demonstrate the strength of the influence that independent variables have on a dependent variable. In Table 15 it is revealed that Poverty% and PerPupil\$ were the only statistically significant variables ( $p = .012$  and  $p = .049$  respectively). A/SRatio ( $p = .926$ ),

Disability% ( $p = .807$ ) and LEP% ( $p = .752$ ) did not reveal a statistically significant influence on EFFECT%. Poverty% had a standardized beta = .187 demonstrating a weak positive influence on EFFECT%, and PerPupil\$'s standardized beta was a positive .125, showing a weaker influence on EFFECT%. With a model  $R$  squared value of 6.3%, and weak beta values for Poverty% and PerPupil\$ there was no logical reason to analyze the data any further.

Table 14

*Effective Percentage Coefficients Table*

Model		Unstandardized coefficients		Standardized coefficients			Collinearity statistics	
		B	Std. error	$\beta$	$t$	$p$	Tolerance	VIF
1	(Constant)	65.673	8.973		7.319	.000		
	A/SRatio	-.001	.007	.006	-.094	.926	.989	1.011
	Disability%	-.128	.159	.052	.807	.420	.943	1.060
	Poverty%	.108	.043	.187	-2.521	.012	.718	1.394
	LEP%	-.047	.148	-.024	-.317	.752	.703	1.422
	PerPupil\$	.001	.000	.125	1.978	.049	.990	1.010

Dependent Variable: EFFECT%

I ran the second model using all five predictor variables with the percentage of teachers scored highly effective as the dependent variable. The model summary is shown below (Table 15) with the  $R$  squared revealed as .062.

Table 15

*Percentage of Teachers Scored Highly Effective Model Summary*

Model	<i>R</i>	<i>R</i> <sup>2</sup>	Adjusted <i>R</i> <sup>2</sup>	Std. error of the estimate
1	.251 <sup>a</sup>	.062	.042	13.94036

*Note.* Predictors: (Constant), PerPupil\$, LEP%, A/SRatio, Disability%, Poverty%<sub>a</sub>

The ANOVA data (Table 16) shows that the model was statistically significant with  $F = 3.099$ ,  $df = 5$ , and  $p = .010 < .05$ .

Table 16

*Percentage of Teachers Scored Highly Effective ANOVA Table*

Model		Sum of squares	<i>df</i>	Mean square	<i>F</i>	<i>p</i>
1	Regression	3010.832	5	602.166	3.099	.010 <sup>b</sup>
	Residual	44586.711	236	194.334		
	Total	48873.541	241			

*Note.* Dependent variable: HIGH%<sub>a</sub>  
 Predictors: (Constant), PerPupil\$, LEP%, A/SRatio, Disability%, Poverty%<sub>b</sub>

The coefficients table (Table 17) revealed that Poverty% was the only statistically significant variable ( $p = .011$ ) and a weak negative influence ( $\beta = -.190$ ). PerPupil\$ had a significance value of .067, close to  $p < .05$  and had a positive beta value of .116 showing a weak influence on EFFECT%. With just one significant variable (at the  $p < .05$  threshold) and a model *R* squared value of 6.2%, there was no logical reason to analyze the data any further.

Table 17

*Highly Effective Percentage Coefficients Table*

Model		Unstandardized		Standardized		Collinearity		
		coefficients		coefficients		statistics		
		B	Std. error	$\beta$	<i>t</i>	<i>p</i>	Tolerance	VIF
1	(Constant)	35.013	8.958		3.909	.000		
	A/SRatio	4.851	.007	.000	.007	.995	.989	1.011
	Disability%	.139	.159	.057	.877	.381	.943	1.060
	Poverty%	-.109	.043	-.190	-2.548	.011	.718	1.394
	LEP%	-.038	.148	-.020	-.260	.795	.703	1.422
	PerPupil\$	.001	.000	-.116	1.838	.067	.990	1.010

*Note.* Dependent variable: HIGH%.

Using the 242 schools in the sample, the results suggest that the predictor variables do not influence the percentages of teachers that are rated effective or highly effective by their evaluating administrators, and, specifically, administrator/student ratios are not related nor can they be used to predict the percentage of teachers in a school who will be rated effective and highly effective on their summative evaluations. The only variables that showed any statistical significance was the influence of poverty% and PerPupil\$ on EFFECT%, yet the *R* squared for the model was just .063. When the influence on HIGH% was measured, the model demonstrated an *R* squared of .062 with only Poverty% being a statistically significant variable, and PerPupil\$

having a significance of .067. In both cases the influence was minimal, but it does suggest that poverty decreases the percentage of teachers being rated as highly effective.

**Research Question 1:** What is the influence of school administrator/student ratios at elementary schools containing grades PreK–6 (as identified by containing three consecutive grades or more, with no class above the sixth grade) on the percentage of teachers being rated effective on their summative evaluations scored by the administrators at those schools?

**Null Hypothesis 1:** No statistically significant relationship exists between school administrator/student ratio and the percentage of teachers scored effective by administrators at those schools.

The null hypothesis was retained. There was no statistical relationship between administrator/student ratio and the percentage of teachers scored effective by evaluating administrators at those schools.

**Research Question 2:** What is the influence of school administrator/student ratios at elementary schools containing grades PreK–6 (as identified by containing three consecutive grades or more, with no school class above the sixth grade) on percentage of teachers being rated highly effective on their summative evaluations scored by the administrators at those schools?

**Null Hypothesis 2:** No statistically significant relationship exists between school administrator/student ratios and the percentage of teachers scored highly effective by administrators at those schools.

The null hypothesis was retained. There was no statistical relationship between administrator/student ratio and the percentage of teachers scored highly effective by evaluating administrators at those schools.

Overall it was determined that the school and staff variables studied did not have statistically significant influence on summative ratings teachers receive evaluated by their building administrators. Specifically, administrator/student ratios could not be used to predict the percentages of teachers rated effective or highly effective.

## CHAPTER V CONCLUSIONS AND RECOMMENDATIONS

In this chapter I present conclusions as well as recommendations for policy, practice, and future research. Using my sample of 242 elementary schools (K–6) it is evidenced that the only factors that can predict the difference in scores between the percentage of teachers that receive effective and highly effective ratings is the percentage of students living in poverty and per-pupil expenditures (only in the EFFECT% model with a  $p = .049$ ); however, the models only represent 6.3% and 6.2% of the variance and cannot be considered a substantial statistic PerPupil\$ also had a  $p = .049$ . There was no statistically significant relationship or predictive nature between administrator/student ratio and teacher evaluation scores.

### **Conclusions**

Of the 242 schools that met the criteria to be included in my sample, it was demonstrated that teachers in schools with high administrator/student ratios, high percentages of students with an IEP, or that are categorized as Limited English Proficient have not had their summative evaluation scores biased because of those factors. However, poverty and per-pupil expenditure do have a statistically significant influence on the summative evaluation scores that teachers receive, albeit weak. It is worth noting, for people that work in schools with impoverished student populations, that poverty even has a small influence on how teachers are scored on their summative evaluations.

High administrator/student ratios can influence the available time administrators have in the daily operation of their schools, as almost all of the essential administrator tasks identified in the literature have to do with relationships with people (Donaldson, 2011; Grissom & Loeb, 2011; Waters, Marzano, & McNulty, 2003): the more students, teachers, parents, and community members, the more time that has to be exhausted. Waters et al., in a meta-analysis of 254 studies,

highlighted the complexity and myriad of duties that principals must perform to be successful. One of their findings was that principals' time is best spent in classrooms and instructional improvement. Using survey research Grissom and Loeb (2011) studied 314 principals (and faculty at their schools) from Miami-Dade County (FL) public schools with the purpose of being able to come to conclusions concerning the influence of principal time-management types and on school-level outcomes (including teacher opinions). They studied 42 principal tasks, related to 21 responsibilities, and defined five types of administrator time: (a) instructional management, (b) internal relations, (c) organizational management, (d) administration, and (e) external relations. They definitively found that different time-management styles impact schools in different ways. Donaldson (2011) qualitatively studied 30 principals and found that principals felt that their ability to effectively evaluate teachers was hindered by "a limited opportunity to observe and document representative teaching" (p. 17). High administrator/student ratios impact an administrator's available time; however, the results from this study suggest that the ratios in this sample did not substantially influence the summative evaluation ratings that teachers received from their evaluating administrators. The fact that administrator/student ratios did not influence teacher evaluations scores can demonstrate that administrators are able to prioritize their duties while adhering to TEACHNJ's mandates.

In some cases, schools with high administrator/student ratios may be the same schools that have low per-pupil expenditures, although the results from the correlation matrix do not suggest that is the case in this sample. Low budgetary spending can create issues such as a lack of textbooks and materials, low teacher salaries, and limited availability of extra-curricular activities. Parrish, Matsumoto, and Fowler (1995), writing for the National Center for Education Statistics, found that there is a connection between districts with students of low SES and low

per-pupil expenditures. Stull (2013) studied SES (using 900 kindergarten programs as a sample) and parental expectations and found that families of high SES use their means in multiple ways to help their children reach higher achievement levels. In New Jersey there is a 100% correlation between SES and student test scores, measured using DFG (NJDOE, 2004), and while per-pupil expenditure is not exactly SES, I found that there was a slight statistical connection between high PerPupil\$ and a greater percentage of teachers being scored highly effective as opposed to effective.

The results from this study suggest that poverty (SES) influences teacher summative evaluation scores in the direction of EFFECT% versus HIGH%, perhaps dampening ratings. The *Equality of Educational Opportunity* (Coleman et al., 1966) study was the initial and foremost study that examined SES. The authors demonstrated that the factors that make up the characteristics of the life of poor minority children negatively impact their achievement levels. This has been backed up by numerous large-scale studies in the 50 years that have followed (Caldas & Bankston, 2001; Kozol, 1991; Sirin, 2005; Tienken, 2012). Abraham Maslow established that certain basic needs have to be met in order for human beings to naturally develop and acquire a sense of self-esteem and love in their environments (Thomas, 2005). These basic physiological and safety needs are outlined in Maslow's hierarchy of needs and include the need for food, sleep, health, employment, safety, monetary resources, property, and reproduction. Humanistic theory suggests that learning is very personal phenomena that occurs when individual goals are set. When basic needs are not met, then one's personal goals cannot reach beyond attempting to meet those survival needs (Buhler & Massarik, 1968). In my study, Poverty% was negatively correlated with teachers receiving a highly effective rating.

Poverty%'s influence predicted that teachers of students with low SES would have a higher rate of Effect% compared to High%.

### **Recommendation for Policy**

Both regression model summaries revealed an *R* squared of just 6.3% ( $\beta = .187$ ) and 6.2% ( $\beta = -.190$ ). Furthermore, the preliminary descriptive research demonstrated that for all intents and purposes teachers are not being evaluated as ineffective or partially effective in New Jersey. Furthermore, receiving an ineffective or partially effective score only puts teachers on corrective action plans and simply starts the process of possible tenure charges. Out of the 242 schools in my study, only six schools reported two teachers as either partially effective or ineffective, and 30 schools reported just one.

The New Jersey Department of Education's Executive Summary (NJDOE, 2004), detailing data on District Factor Groups (DFG), the system used to categorize school districts based on SES after the 1990 and 2000 censuses, revealed a 100% correlation between school district SES and student achievement levels. The lower the SES levels the lower the average test scores for that district. Still pertinent today, these data can be found on the state website under the Data tab (NJDOE, 2004).

Leading up to the passing of the TEACHNJ, under the direction of Governor Chris Christie, the state department of education created an Education Transformation Task Force that produced a 239-page report. A justification for the over 40 recommendations made was the achievement gap that existed in New Jersey. In the report it is written (Hespe et al., 2014):

When large, predictable swaths of our young people — low-income boys and girls, African Americans, Hispanics and others — consistently fall behind, we must honestly acknowledge that our public schools are not delivering on their promise and purpose. Poverty and other forces matter, but so do our schools. Good schools help students transcend the challenges they face, while mediocre ones are unequal to the task. Yes, there are enormous challenges associated with enabling every child to succeed. But that

doesn't take away the responsibility of adults in our public education system to organize their efforts, deploy their resources, and shape their craft to advance student learning to consistently higher levels. For decades, many of the State's education leaders have shown an extraordinary commitment to New Jersey's children, working to change policies and practices to drastically improve achievement. But a clear-eyed accounting of our standing forces us to admit that the current order has not produced the results we so desperately need. Results from the New Jersey Assessment of and Skills and Knowledge (NJ ASK) reveals that 40 percent of third graders in New Jersey are not reading at grade level. According to the National Assessment of Educational Progress (NAEP), only a handful of states have a larger achievement gap between economically disadvantaged students and their wealthier peers. Nearly 90 percent of students entering some of our community colleges require remediation in reading, writing or math. (pp. 4–5)

Almost simultaneously to the Task Force's report, the Christie Administration cut every school district's state aid by 5% of its total budget. For example, if a district's budget was \$1,000,000 and of that budget \$200,000 came from state aid, then that district's state aid was reduced by \$50,000 as that is 5% of its total budget. State aid is almost solely based on the percentage of students living in poverty, which meant that some wealthy districts did not have 5% of their budget cut because they received less than 5% of their total budget from the state. For example, if a district has a total budget of \$1,000,000 and \$20,000 of their budget comes from the state, they only have 2% of their budget cut, because it only receives \$20,000 from the state.

“AchieveNJ: Increasing Student Achievement through Educator Effectiveness” (NJDOE, 2015a, slide 1) is one of many publicly available guides to the TEACHNJ law. The state also advertised that the bill was created to make it easier to fire underperforming teachers, “*streamlining the tenure arbitration system for all staff members*” (NJDE, n.d.d, p. 4). Removing ineffective teachers from our poverty-stricken districts seemed to be New Jersey's new solution to shrink the achievement gap.

In the aforementioned *Equality of Educational Opportunity* report (Coleman et al., 1966), the authors used production function theory to demonstrate that student achievement and poverty are interwoven together, and the evidence still holds strong today. In my research I found a very

small connection between poverty and the percentages of teachers that were scored effective as opposed to highly effective (and the scoring of teachers as partially effective/ineffective almost does not happen at all).

Knowing the irrefutable connection between poverty and low student achievement, a full reexamination of education policy is needed. The factors that New Jersey uses to measure SES are “1) Percent of adults with no high school diploma 2) Percent of adults with some college education 3) Occupational status 4) Unemployment rate 5) Percent of individuals in poverty 6) Median family income” (NJDE, 2004, p. 1). The TEACHNJ act of 2012 was implemented to help solve the achievement gap and help students overcome the issues they are faced with because of their low SES factors. Policy writers need to ask themselves if easing the process of filing tenure charges against public school teachers, is really the answer.

When Socrates described his philosophy of education to Plato, he stated that knowledge (learning) had to be a universal and ethical right and that societal strength is demonstrated when all people, from all social classes, have the freedom to pursue their interests and personal talents. Ecological human development theorist Bronfenbrenner (1979) explained that children develop through their interpersonal relations and experiences within given settings called microsystems. The child’s home (including the immediate neighborhood), school, and peer group are microsystems that immediately impact development (Thomas, 2005).

Explaining how the settings of a child’s microsystems impact his or her development, Thomas (2005) wrote, “The environment also initiates transactions with the child that either promote or thwart development,” and that, “Influences on development can also be instigated by such physical conditions of environment as the size and furnishings of the child’s dwelling, the child’s diet, the quality of the air, the surrounding noise, threats of danger and far more” (p. 352).

As the child develops within immediate essential settings he or she continues to develop as interactions occur with larger collections of microsystems under an entire macrosystem, which represents the evolution of his or her culture (e.g., ethnicity and government within a country).

Like Bronfenbrenner and Maslow, Lawrence Kohlberg believed that human interactions with environmental factors impact the development of children (Thomas, 2005). Kohlberg measured development in terms of moral growth and created a 6-part scale of moral development. In his research (using mostly adolescent boys as his sample) he asserted that low level moral development is viewing the world in terms of what is right and wrong and in how actions impact oneself.

Higher moral development starts when people are able to put themselves in others' shoes (Golden Rule) and view all people as equals. Decisions and thoughts are measured by how any human being would be impacted in a particular situation. The sixth and highest stage of moral development occurs when people are ready to question society and bring about change. An example of reaching the highest stage of moral development (called the universal ethical principle orientation) is when people seek to bring about change in a society when property supersedes life and the basic needs of any individual.

Ecological and humanistic theories elucidate how important the immediate environment is to a child's natural development. Humanism theory suggests that, in order to healthily develop, children need to have basic needs met. Moral development points to a moral society as a one that is ethically driven to universally protect all people as equal human beings. My research demonstrated that there is little connection between identifying teacher effectiveness (or a lack thereof), and identifying teacher effectiveness was supposed to help create achievement gains for poor children. Moreover, theory suggests that even if we were able to identify and dismiss

ineffective teachers, that would not address any of the real poverty issues that impact how children interact with their educational microsystem, the issues that are the proven reason for the achievement gap. Actions have to be about measures that can help a child overcome the poverty conditions that he or she is forced to interact with in the world outside of a school building.

The federal government recently instituted a tax cut that helps the wealthy, and whether one believes it will eventually trickle down to the poor and middle class, it is overwhelmingly accepted that it is biased towards the rich (Bump, 2017; Hoxie, 2017). Poverty outside of the educational biosphere is not being addressed as a societal issue, but that does not mean that education policy cannot address poverty.

The one measure that has been proven to help poor children, where choice, acceptance, and self-worth can be managed and fostered, is CSR. The Tennessee STAR, CSR experiment (with an experimental group of 13–17 students per classroom, compared to a full-class of 22–25 students and a full-class with a teacher aide), had a sample of over 6,000 students in 329 classrooms and demonstrated that CSR in K–3 classrooms had lasting and positive impacts on the achievement of children from all demographics (Achilles, Finn, & Bain, 1997; Mosteller, 1995; Tienken & Achilles, 2009). Furthermore, fewer grade-level retentions, higher high school graduation rates, an increase in taking advanced classes in high school, and more college entrance examinations occurred (Finn & Achilles, 1999). Finn, Panno, and Achilles (2003) posited that one major reason why CSR works is that it allows for greater student engagement, an essential learning behavior that, in the Tennessee STAR study, was only attributed to the low number of students in classrooms, not because of any measure of teacher quality. An essential component to the Tennessee STAR study was that teachers were assigned randomly, and teacher quality was not a part of the equation; yet, “The benefits were substantially greater for minority

students or students attending inner-city schools in each year of the study” (Finn & Achilles, 1999, p. 98).

In “The Purloined Letter” (a story that has been dissected and analyzed by scholars far and wide), Edgar Allan Poe (1844) posited this about solving problems:

“There is a game of puzzles,” he resumed, “which is played upon a map. One party playing requires another to find a given word—the name of town, river, state or empire—any word, in short, upon the motley and perplexed surface of the chart. A novice in the game generally seeks to embarrass his opponents by giving them the most minutely lettered names; but the adept selects such words as stretch, in large characters, from one end of the chart to the other. These, like the overly large lettered signs and placards of the street, escape observation by dint of being excessively obvious; and here the physical oversight is precisely analogous with the moral inapprehension by which the intellect suffers to pass unnoticed those considerations which are too obtrusively and too palpably self-evident...” (p. 17)

It is right there in front of our faces written clearly across the entire blueprint of educational reform: Policy needs to be based in proven research.

### **Recommendation for Practice**

Highly effective ratings can be used as a powerful tool to build efficacy, create positive emotions, and empower teachers of students with an IEP, LEP, and/or living in poverty.

Danielson’s Framework for Teaching rubric, the most commonly used evaluation framework in New Jersey, has 22 components and 76 elements that principals can use when evaluating their teachers. This means that there are 76 different ways a principal can choose to empower his or her teachers and boost the morale of those that teach the neediest children.

IDEA (1991) and NCLB (2002) mandated that children with disabilities and those that speak a first language other than English (respectively) take the same standardized test as typical students. The results from this study pointed to the fact that there is no positive connection between teaching students of special populations (disabilities, LEP, low SES) and a teacher being rated as highly effective. The potential negative influences of having a cognitive disability

or not speaking English are obvious in terms of results on standardized tests, and “Poor children in the United States start school at a disadvantage in terms of their early skills behavior and health” (Isaacs, 2012, P 1.). Elementary school teachers may be the only people that have the opportunity to help fill in the basic educational and human needs that children with disabilities, language barriers, and poverty (sometimes with two or all three of these factors) have. Jensen (2013) described how teachers have to fill in the emotional needs that children living in poverty often do not receive at their homes:

Cooperation, patience, embarrassment, empathy, gratitude, and forgiveness are crucial to a smoothly running complex social environment (like a classroom). When students lack these learned responses, teachers who expect humility or penitence may get a smirk instead, a response that may lead teachers to believe the student has an “attitude.” It’s the primary caregiver’s job to teach the child when and how to display these emotional responses, but when students do not bring these necessary behaviors to school, the school must teach them. (p. 20)

Elementary school students spend an awesome amount of meaningful time with their teachers. Citing Harris (2006), Jensen (2013) identified that all children come to school seeking “The drive for reliable relationships... The strengthening of peer socialization (and)...The quest for importance and social status” (p. 21). Myers and Pianta (2008) found that poor student teacher relationships in kindergarten result in negative outcomes ranging from low grades and test scores (that remain through their educational careers) to an increase in disciplinary infractions and poor work habits. Teachers matter, and motivating the teachers that work with student populations that need emotional support can be a powerful tool for K–6 principals.

Albert Bandura, a human development theorist, researched social learning. He outlined how children learn through the observation of others and described how adults act as models for children, have lasting positive or negative influences on them. Like B.F. Skinner, and behaviorists before him, Bandura believed that a child’s responses to the outside world, and the

reinforcements received, shape the development of that child. He also examined self-efficacy's role in overall human development, which he defined as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). In discussing Bandura's self-efficacy, Thomas (2005) explained that Bandura demonstrated how people's emotions, experiences, and the feedback they receive from their surroundings, impact their motivation for success.

Research has been conducted on teachers' feeling of efficacy. Hoy and Woolfolk (1993) researched the personal self-efficacy of teachers, which they defined as a confidence and belief that he or she can positively influence his or her students' learning. Hoy, Tarter, and Woolfolk Hoy (2006) researched teacher collective efficacy, which builds upon personal efficacy, but is a belief that, together, the entire faculty of a school can make a difference in their students' lives. Academic optimism is a mechanism that uses the promotion of teacher collective efficacy as one of three essential pieces that administrators can promote to positively impact student achievement (McGuigan & Hoy, 2006). Spreading positive self and collective efficacy should be a powerful tool that administrators use; efficacy helps people "to determine how strongly they try to succeed" (Thomas, 2005, p. 157).

The lack of a connection between administrator/student ratios and teacher evaluation scores can be a call for school principals to change how they use evaluation models. I recommend that they be used as tools to promote a positive school climate. Tziner, Murphy, and Cleveland (2001) examined the relationship between raters' (in education that would be evaluators) attitudes toward their organizations (workplace), beliefs about the appraisal systems used to evaluate subordinates, and confidence in the systems and resulting rating behaviors. The study was built on self-efficacy theory and research on the influence that organizational climate

has on an individual's performance. Citing Steers and Lee (1983) Tziner et al wrote, "A participative climate is characterized by cooperative relationships, individual responsibility, trust, and communication. All these factors are likely to reduce supervisor-subordinate conflict" (p. 227).

Tziner et al. (2001) used data from seven different organizations in three countries (the USA, Canada and Israel); each organization was uniquely different from the others. The types were of: (a) military cadets (USA), (b) workers in a utility company (USA), (c) students at the University of Montreal, (d), members of the Association of Managers and Professionals of the University of Montreal, (e) managers at an aluminum manufacturing plant, and (f & g) various managers throughout Israel.

Rensis Likert (known most for the widely used Likert survey method that measures people's responses to questions on five levels of agreement) was an organizational psychologist that developed the theory of participative management. Likert (1961) demonstrated that "managerial climate affects the behavior of industrial work groups (as cited in Kaczka & Kirk, 1968, p. 254). There is a substantial amount of research that identifies the power that administrators have in promoting a positive school climate (Hallinger & Heck, 1996; Leithwood & Jantzi, 2000; Witziers, Bosker, & Kruger, 2003). Halverson and Clifford (2006) found that a challenge to successful use of teacher evaluation lied in the fact that evaluators were concerned with pleasing their teachers in the process.

In the lead-up to my study I found that teachers in New Jersey are not being evaluated as ineffective or partially effective, and the school and student factors I measured do not have a statistically significant impact on whether teachers receive effective or highly effective ratings. TEACHNJ mandates that all results to be reported to the state Department of Education (Ch. 26;

C.18A:6-117); however, the reporting of data does not seem to be providing what state officials sought. Instead of using evaluation models, and all the components they incorporate (Danielson has 76 elements), as reporting mechanisms they should be used as tools to create a positive school climate.

Principals in New Jersey have been using the new evaluation rubric for 4 years. The confidence that comes with experience should be used as a way to facilitate positive relationships with faculty. When a principal shows trust and acceptance of an evaluation model, he or she can convey that assuredness to faculty. Tziner et al. (2001) demonstrated that “confidence in appraisal systems seems to be the single most important predictor” (p. 276) for fostering a positive organizational climate.

### **Recommendations for Future Research**

I was able to glean three clear avenues for future research. Summative teacher evaluation scores have not been used as a dependent variable in any prior research. As independent variables I used the available public data deemed pertinent and published by the New Jersey Department of Education. Therefore, the first area for future examination is to measure other factors that influence the evaluation process. With these data I only included 242 specific K–6 elementary schools. Middle school administrators and faculty are uniquely different from those at the elementary school level. Middle school teachers must hold a degree and certificate in their specific subject taught (math, science, language arts, etc.), and teachers’ levels of education were not used in my study. Greenlee (2007) elucidated that middle school budget concerns are entirely different from elementary school expenditures. How does the publicly available secondary education data, including teachers’ level of education, influence teacher summative scores evaluated by their building administrators?

Within this same bracket of recommended research, the actual evaluation programs used by each individual district could be an important independent variable to examine in the future. As pointed out in Chapter I, there are 26 different approved programs (NJDOE, 2015a) used by the more than 500 school districts in New Jersey. Sixty percent of New Jersey's school districts use an edition of Charlotte Danielson's Framework for Teaching (Mooney, 2013), and the other four most commonly used programs are James Stronge's Teacher and Leader Effectiveness Performance System, the Mid-Continent Research for Education and Learning (McREL), Marzano's Causal Teacher Evaluation Model, and the Marshall Rubrics (Schoenlank, 2017). An analysis to determine how these programs influence the effectiveness scores teachers receive from their evaluating administrators could help in the differentiation between teachers that are rated effective and highly effective. Further, if any program can predict teacher scores in any direction that could then provide useful information for further quantitative and qualitative dissection.

The second path for future research would be to analyze the connection between the purposes of TEACHNJ's four-tiered evaluation rubric and the intended results of the law. One purpose of the law was to streamline the tenure process; ergo, there should be a connection between the institution of the new evaluation system and the identification and dismissal of ineffective teachers. I have already pointed out that a quick analysis of the data demonstrates that is not happening. However, the other stated purpose is to improve achievement, so, even if more ineffective teachers have not lost their tenure, one could argue that the process of identifying teacher effectiveness will boost overall achievement in the state, and more importantly that of our minority students living in poverty. In my research, I only found an *R* squared value of .063 concerning the influence of poverty and percentage of teachers that received effective evaluation

scores when compared to highly effective. However, even if my coefficient of determination was greater than 50% (in connection to either effective or highly effective scores), studying how using the effectiveness scale influences academic achievement needs to happen.

Thirdly, a qualitative examination of teacher perceptions and reactions of what it is like being evaluated under the guidelines of TEACHNJ should be undertaken. Tuytens and Devos (2014) found that when teachers view feedback from their evaluators as useful they go on to seek out relevant opportunities for professional growth. Furthermore, assessment feedback that is not aligned with a teacher's belief about his or her classroom performance leads to a decrease in his or her sense of personal self-efficacy (Ham, Duyar, & Gumus, 2015). How do teachers feel that the evaluation provisions of TEACHNJ have influenced their daily interactions in their schools?

The TEACHNJ Act of 2012 was the dilemma that was the motivation for the derivation of my research problem. The law was intended to increase student achievement through the identification of teacher effectiveness, while citing that teacher effectiveness was just one of many factors that affect achievement. My study was the first of its kind to examine teacher summative evaluation scores as a dependent variable. A continuation of using teacher evaluation as a production function output should be undertaken with the importance it has been given under the law. If it is continued to be found that school, student, and teacher variables do not influence evaluation scores, then an investigation that explains why should be considered. Moreover, if research continues to prove unable to glean tangible data from teacher evaluations scores under TEACHNJ's regulations, then all stakeholders in the education of our children should consider making consequential decisions based in proven research, and not in a law that is focused more on the filing of tenure charges than anything else.

## References

- Achilles, C. M. (2012). Class-size policy: The STAR experiment and related class-size studies (NCPEA Policy Brief, Vol 1., No. 2) Retrieved from <http://www.nceapublications.org>
- Achilles, C. M., Finn, J. D., & Bain, H. P. (1997). Using class size to reduce the equity gap. *Educational Leadership*, 55(4), 40–43.
- Aldeman, C. (2017). The teacher evaluation revamp, in hindsight. *Education Next*, 17(2), 60–68.
- American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115, 181. (2009).
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., et al. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.
- Baker, B. D., Oluwole J., & Green P.C., III (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the race-to-the-top era. *Educational Policy Analysis Archives*, 21(5), 1–71.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.
- Bass, M. B. & Riggio, E. G. (2006). *Transformational Leadership (2<sup>nd</sup> ed.)*. Mahwah, NJ: Lawrence Elbraum Associates.
- Beck, L.G., & Murphy, J. (1993). *Understanding the principalship: Metaphorical themes 1920s–1990s*. New York: Teachers College Press.
- Belfield, C., & Heywood, J. (2008). Performance pay for teachers: Determinants and consequences. *Economics of Education Review*, 27(1), 243–252.
- Berndt, E., & Christenensen, H. B. (1973). The internal structure of functional relationships: Separability, substitution and aggregation. *Review of Economic Studies*, 40(3), 403–410.
- Betebenner, D. W. (2009). Norm and criterion-referenced student growth. *Education Growth and Measurement: Issues and Practice*, 28(4), 42–51.
- Betebenner D., W. (2011) *A technical overview of the Student Growth Percentile methodology: Student Growth Percentiles and percentile growth projections/trajectories*. The National Center for the Improvement of Educational Assessment. Retrieved from [http://www.nj.gov/education/njsmart/performance/SGP\\_Technical\\_Overview.pdf](http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf)

- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3–15.
- Borg, M. O. & Shapiro, S. L. (1996). Personality type and student performance in principles of economics. *Journal of Economic Education*, 27(1), 3-25.
- Bowles, S. (1970). Towards and educational production function. In Hansen, L. W. Education, income and human capital (11-70). Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Buhler, C., & Massarik, F. (1968). *The course of human life: A study of goals in the humanistic perspective*. Oxford, England: Springer.
- Bump, P. (2017, December, 23). What to look for from the Republican tax bill, month by month. *Washington Post*. Retrieved from [https://www.washingtonpost.com/news/politics/wp/2017/12/22/what-to-look-for-from-the-republican-tax-bill-month-by-month/?utm\\_term=.4e42af85e4fc](https://www.washingtonpost.com/news/politics/wp/2017/12/22/what-to-look-for-from-the-republican-tax-bill-month-by-month/?utm_term=.4e42af85e4fc)
- Caldas, S. J., & Bankston III, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *Journal of Educational Research*, 90(5), 269.
- Callahan, K., & Sadeghi, L. (2012-2013). *New Jersey teacher tenure reform* (White Paper Series). Kean University Center for History, Politics and Policy. Retrieved from <http://chpp.kean.edu/white-paper/new-jersey-teacher-tenure-reform>
- Castellano, K. E., & Ho, A. D. (2013). Contrasting OLS and quantile regression approaches to student "Growth" percentiles. *Journal of Educational and Behavioral Statistics*, 38(2), 192–215.
- Cawelti, G. (1982). Training for effective school administrators. *Educational Leadership*, 39(5), 324–329.
- Cerf, C. D. (2012). *Education funding report*. Retrieved from <http://www.nj.gov/education/stateaid/1213/report.pdf>
- Coleman, J. S. et al. (1966). *Equality of educational opportunity*. Washington, DC: National Center for Education Statistics.
- Colorado Department of Education. (n.d.). *Rubric for evaluating Colorado teachers*. Retrieved from <http://www.cde.state.co.us/sites/default/files/TeacherRubric.pdf>
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: ASCD.

- Danielson, C. (2010). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35–39.
- Danielson, C. (2015). Framing discussions about teaching: A streamlined version of the widely used Framework for Teaching can empower professional conversations. *Educational Leadership*, 72(7), 38–41.
- Darling-Hammond, L. (1990). *Teacher evaluation in transition: Emerging roles and evolving methods*. In Millman, J. & Darling-Hammond, L. The new handbook for teacher evaluation (17-35). Newbury Park, CA. Sage.
- Darling-Hammond, L., Wise, A. E. & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 258–328.
- Donaldson, M. L. (2011). Principals' approaches to Developing teacher quality: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Center for American Progress*. Retrieved from <https://files.eric.ed.gov/fulltext/ED535861.pdf>
- Douglas, S. & Sulock, J. (1995). Estimating educational production functions with correction for drops. *Journal of Economic Education*, 26(2), 101-112.
- Ehlert, M., Koedel, C., Parsons, E., & Podgursky (2013). The sensitivity of value-added estimated to specification adjustments: Evidence from school-and teacher-level models in Missouri. *Journal of Statistics and Public Policy*, 1(1), 19–27.
- Evans, B. R., Willis, F., & Moretti, M. (2015). Editor and section editor's perspective article: A look at the Danielson Framework for Teacher Evaluation. *Journal of the National Association for Alternative Certification*, 12(1), 21–26.
- Farhat, A. (2016). Professional development through clinical supervision. *Education*, 136(4), 428–436.
- Feelhaver, C. T. (1927). The duties of high-school principals in the state of Nebraska. *The School Review*, 35(3), 188–193.
- Field, A. P. (2013). *Discovering statistics using SPSS* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Finn, J. D., & Achilles C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27(3), 557–577.
- Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21(2), 97–109.

- Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The 'why's' of class size: Student behavior in small classes. *Review of Educational Research*, 73(3), 321–368.
- Furney, K., Aiken, J., Hasazi, S., & Clark-Keefe, K. (2005). Meeting the needs of all students: Contributions of effective school leaders. *Journal of School Leadership*, 15(5), 546–570.
- Gage, N., & Needels, M. (1989). Process-product research on teaching: A review of criticisms. *The elementary school journal*, 89(3), 253–300.
- Gius, M. (2014). Using fixed effects to estimate the impact of merit pay on teacher job satisfaction. *Journal of Economics and Economic Education Research*, 15(1), 17–30.
- Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2004). *Supervision and instructional leadership: A developmental approach* (6th ed.). Boston, MA: Pearson Education and Allyn and Bacon.
- Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, 44(2), 87–95.
- Goldhaber, D. D. & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human resources*, 32(3), 505-523.
- Gordon D. & Vaughn, R. (2011). The historical role of production function in economics and business. *American Journal of Business Education*, 4(4), 25–29.
- Greenlee, B. J. (2007). When school advisory councils decide: Spending choices for school improvement. *Planning and Changing*, 38(3), 222–244.
- Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361-396
- Grissom, J. A., & Loeb, J. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal* 48(5), 1091-1123.
- Grissom, J. A., Loeb, S., & Mitani, H. (2013). Principal time management skills: Explaining patterns in principals' time use, job stress, and perceived effectiveness. *Journal of Educational Administration*, 56(3), 773–793.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, 33(3), 329-352.
- Hallinger, P. (2011). Leadership for learning: lessons from 40 years of empirical research, *Journal of Educational Administration*, 49(2), 125-142.

- Hallinger, P., & Heck, R. H. (1996) Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, 32(1), 5–44.
- Halverson, R. R. & Clifford, M. A. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. *Education Administration Quarterly*, 42(4), 578-619.
- Ham, S., Duyar, I., & Gumus, S. (2015). Agreement of self-other perceptions matters: Analyzing the effectiveness of principal leadership through multi-source assessment. *Australian Journal of Education*, 59(3), 225–246.
- Hanushek, E. A. (2005). Why quality matters in education. *Finance and Development*, 42(2), 15-19.
- Hanushek, E. A. (2010). Education production functions: Developed country evidence. In *International encyclopedia of education*. (Vol. 2, pp. 407–411). Oxford, England: Elsevier.
- Hanushek, E. A. & Kain, J. F. (1972) On the value of equality of educational opportunity as a guide to public policy. In Mosteller, F. & Moynihan, D. P. On equality of educational opportunity, (116-146). New York, NY: Random House.
- Hespe, H. et al. (2012). *Education transformation task force: Final report* (New Jersey Department of Education). Retrieved from <http://www.nj.gov/education/reform/ETTFFinalReport.pdf>
- Hogan, R. L. (2007). The historical development of program evaluation: Exploring the past and present. *Online Journal of Workforce Education and Development*, 2(4), 1–14. Retrieved from <http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1056&context=ojwed>
- Hoppey D., & McLesky J. (2013). A case study of principal leadership in an effective inclusive school. *Journal of Special Education*, 46(4), 245–256.
- Hoxie, J. (2017, November, 3). Trump’s tax cuts are the biggest wealth grab in modern history. *Fortune*. Retrieved from <http://fortune.com/2017/11/03/trump-gop-tax-plan-cuts-2017/>
- Hoy, W. & Miskel, C. (2008). *Educational Administration: Theory, Research and Practice*. New York: McGraw-Hill
- Hoy, W., & Woolfolk, A. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal*, 93(4), 355-372.

- Hoy, W. K., Tarter, C. J., & Woolfolk Hoy, A. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, 43(3), 425–446.
- Hunter, M. (1980). Six types of supervisory conferences. *Educational Leadership*, 37(5), 408.
- Hunter, M. (1984). Knowing, teaching and supervising. In P. Hosford (Ed.), *Using what we know about teaching* (169-193). Alexandria, VA: ASCD.
- Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (1991).
- Isaacs, J.B. (2012). Starting a school at a disadvantage: The school readiness of poor children. *The Social Genome Project, The Brookings Institute*. Retrieved from <https://pdfs.semanticscholar.org/106b/87b07e4dd399a52e9d473343df2fe95d8fce.pdf>
- Jensen, E. (2013). *How poverty affects classroom engagement. Educational leadership*. Alexandria, VA: ASCD.
- Jia, S., Long, Q., Wang, R. Y., Yan, J., & Kang, D. (2016). On the inapplicability of the Cobb-Douglas production function for estimating the benefit of water use and the value of water resources. *Water Resources Management*, 30(10), 3645–3650.
- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, 30(2), 3–13.
- Jordan, H. R., Mendro, R. L., & Weerasinghe, D. (1997). Teacher effects on longitudinal student achievement: A preliminary report on research on teacher effectiveness. Paper presented at the National Evaluation Institute, Indianapolis, IN.
- Kaczka, E., & Kirk, R. (1968). Managerial climate, work groups, and organizational performance. *Administrative Science Quarterly*, 12, 253–272.
- Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1), 52–54.
- Klein, A. (2015). *ESEA reauthorization: The Every Student Exceeds Act explained* [Web log post]. Retrieved from [http://blogs.edweek.org/edweek/campaign-k-12/2015/11/esea\\_reauthorization\\_the\\_every.html](http://blogs.edweek.org/edweek/campaign-k-12/2015/11/esea_reauthorization_the_every.html)
- Klump, R., McAdam, P. & Willman, A. (2012). The normalized CES production function: Theory and empirics. *Journal of economic surveys*, 26(5), 769-799.
- Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York, NY: Crown.
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2015). Factorial ANOVA and ANCOVA (Chapter 9). In *IBM SPSS for Intermediate Statistics: Use and Interpretation* (5th edition, pp. 188–212). New York: Routledge.

- Leithwood, J., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, 38(2), 112–129.
- Mansanjala, W. H., & Papageorgiou, C. (2004). The Solow model with CES technology: Nonlinearities and parameter heterogeneity. *Journal of Applied Econometrics*, 19, 171–201.
- Marzano, R. J., Waters, T. & McNulty, B. (2005). *School leadership that works*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mayeske, G. W. et al. (1972). *A study of our nation's schools*. Washington, DC: U.S. Office of Education.
- McGreal, T. (1983). Effective teacher evaluation systems. *Educational Leadership*, 39(4), 303–305.
- McGuigan, L. & Hoy, W. K. (2006) Principal leadership: Creating a culture of academic optimism to improve achievement for all students, *Leadership and Policy in Schools*, 5(3), 203-229.
- McLaughlin, M., & Pfeifer, S. (1988) *Teacher evaluation: Improvement, accountability and effective learning*. New York, NY: Teachers College Press.
- Milanowski, A., & Heneman, H. (2002). *Transforming teacher evaluation: A standards-based teacher evaluation system*. Paper presented at the Sixteenth Annual Conference of the American Evaluation Association. Washington, DC.
- Miles, J. & Shelvin, M. (2001). *Applying regression & Correlation” A guide for students and researchers*. Thousands Oaks, CA: Sage Publications.
- Moody, H. (1968). Plight of the principal. *The Clearing House*, 42(9), 543–545.
- Mooney, J. (2013). New Jersey schools largely choosing the same teacher evaluation model. *The Hechinger Report*. Retrieved from <http://hechingerreport.org/new-jersey-schools-largely-choosing-the-same-teacher-evaluation-model/>
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children: Critical Issues for Children and Youths*, 5(2), 113–127.
- Myers, S., & Pianta, R. (2008). Developmental commentary: Individual and contextual influences on student-teacher relationships and children’s early problem behaviors. *Journal of Clinical Child & Adolescent Psychology*, 37(3), 600–608.

- National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC. Retrieved from <http://www.ed.gov/pubs/NatAtRisk/index.html>
- New Jersey Department of Education. (2004). *District factor groups (DFG) for school districts*. Retrieved from <https://www.state.nj.us/education/finance/rda/dfg.shtml>
- New Jersey Department of Education. (2015a). *AchieveNJ overview presentation—State NJ*. Retrieved from <http://www.state.nj.us/education/AchieveNJ/intro/OverviewPPT.pdf>
- New Jersey Department of Education. (2015b). *Approved teacher practice evaluation instruments*. Retrieved from <http://www.nj.gov/education/AchieveNJ/teacher/approvedlist.pdf>
- New Jersey Department of Education. (n.d.a). *Development of Student Growth Percentiles (SGPs) in New Jersey*. Retrieved from <http://www.nj.gov/education/AchieveNJ/teacher/HighLevelSGPTimeline.pdf>
- New Jersey Department of Education. (n.d.b). *Achieve NJ: Increasing student achievement through educator effectiveness*. Retrieved from <http://www.nj.gov/education/AchieveNJ/intro/OverviewPPT.pdf>
- New Jersey Department of Education. (n.d.c). *Student Growth Objectives*. Retrieved from <http://www.nj.gov/education/AchieveNJ/teacher/SGOOverview.pdf>
- New Jersey Department of Education. (n.d.d). *TEACHNJ guide (updated 6-14)*. Retrieved from <http://www.state.nj.us/education/AchieveNJ/intro/TeachNJGuide.pdf>
- New Jersey Department of Education. (n.d.e). *Staff evaluations 2014-15*. Retrieved from <http://www.state.nj.us/education/data/staff/>
- New Jersey Department of Education. (n.d.f). *2014-2015 school performance reports*. Retrieved from <http://www.nj.gov/education/pr/1415/database.html>
- New Jersey Department of Education. (n.d.g). *Requirements and Resources for Teacher Evaluations in 2014-15*. Retrieved from <http://www.state.nj.us/education/AchieveNJ/resources/TeacherEvaluationRequirementsandResources.pdf>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- Norton, M. S. (1972). Current problems of the high school principal. *The Clearing House*, 46(8), 451–457.

- Olivia, P., & Pawlas, G. (2004). *Supervision for today's schools*. Hoboken, NJ: John Wiley and Sons.
- Paige, R. (2004). *A guide to education and No Child Left Behind*. U.S. Department of Education. Retrieved from [www.ed.gov/nclb/overview/intro/guide/](http://www.ed.gov/nclb/overview/intro/guide/)
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows* (4th ed.) Philadelphia, PA: Open University Press.
- Parrish, T. B., Matsumoto, C. S., & Fowler, W., Jr. (1995). *Disparities in public school district spending: 1989-90*. U.S. Department of Education, National Center for Education Statistics (NCES Research and Development Report No. 95-300). Retrieved from <https://nces.ed.gov/pubs95/web/95300.asp>
- Paxton, P., T. (2016). Everyone agreed on something: How New Jersey's TEACHNJ Act puts testing before students, and firing before teaching. *Seton Hall Legislative Journal*, 40(2), 383-410.
- Peterson, K. (2004) Research on school teacher evaluation. *NASSP Bulletin* 88(639), 60-79.
- Poe, E. A. (1844). The purloined letter. Retrieved from <http://xroads.virginia.edu/~hyper/POE/purloine.html>
- Pogodzinski, B. (2013). Administrative context and novice teacher-mentor interactions. *Journal of Educational Administration*, 53(1), 40-66.
- Powell, A. (2017, October 11). How Sputnik changed U.S. education. *Harvard Gazette: National & World Affairs*. Retrieved from <https://news.harvard.edu/gazette/story/2007/10/how-sputnik-changed-u-s-education/>
- Razik, T. A. & Swanson, A. D. (2001). *Fundamental concepts of educational leadership, 2nd edition*. Upper Saddle River, N.J.: Merrill/Prentice Hall.
- Sammarone, D., R. (2014). *The influence of the length of the school day on the percentage of proficient and advanced proficient scores on the New Jersey Assessment of Skills and Knowledge for grades 6, 7, and 8* (Doctoral dissertation). Retrieved from Proquest Dissertations & Theses @ Seton Hall University. (UMI Number 3639453).
- Sanders, W. L., & Horn, S. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247-256.
- Schoenlank, J. (2017) *School administrators' perceptions of the James Stronge teacher evaluation system* (Doctoral dissertation). Retrieved from Seton Hall University Dissertations and Theses (ETDs). 2429.

- Schundler, B. (2010). *New Jersey Department of Education 2010-2011 state aid*. Retrieved from <http://www.state.nj.us/education/stateaid/1011/CommissionersMemo2011.pdf>
- Schutt, R. K. (2014). *Investigating the social world: The process and practice of research* (8<sup>th</sup> Ed.). Thousand Oaks. CA: Sage.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1), 65–94.
- Stronge, J. H., Holly, B. R., & Catano, N. (2008). *Qualities of effective principals*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Stull, J. C. (2013). Family socioeconomic status, parent expectations, and a child's achievement. *Research in Education*, 90, 53–67.
- TEACHNJ Act of 2012, Ch. 26; C.18A:6-117 (2012).
- Thomas, R. M. (2005). *Comparing theories of child development* (6th ed.). Belmont, CA: Thomson Wadsworth.
- Tienken, C. H. (2012). Poverty matters. *AASA Journal of Scholarship and Practice*, 9(1), 3-6
- Tienken, C. H., & Achilles, C. M. (2006). Making class size work in middle school grades. *AASA Journal of Scholarship & Practice*, 3(1), 26–36.
- Tienken, C. H., & Achilles, C. M. (2009). Relationship between class size and students' opportunity to learn writing in middle school. *Research in the Schools*, 16(1), 13–24.
- Toch, T., & Rothman, R. (2008). Rush to judgment: Teacher evaluation in public education. *Education Sector Reports*. Retrieved from [http://educationpolicy.air.org/sites/default/files/publications/RushToJudgment\\_ES\\_Jan08.pdf](http://educationpolicy.air.org/sites/default/files/publications/RushToJudgment_ES_Jan08.pdf)
- Turnamian, P. G. (2012). *The value of NJ school district demographic data in explaining school district NJ ASK grade 3 language arts and mathematics scores* (Doctoral dissertation). Retrieved from Proquest Dissertations & Theses @ Seton Hall University. (ETDs) 1778.
- Tuytens, M., & Devos, G. (2014). How to activate teachers through teacher evaluation? *School Effectiveness & School Improvement*, 25(4), 509–530.
- Tziner, A., Murphy, K. R., & Cleveland, J. N. (2001). Relationship between attitudes toward organizations and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment*, 9, 226–239.

U.S. Department of Education, Pub. L. No. 111-5, 123 Stat. 115

- van Knippenberg, D. & Sitkin, S. B. (2013). A critical assessment of charismatic—transformational leadership research: Back to the drawing board? *The Academy of Management Annals*, 7(1), 1-60.
- Weisberg, D., Sexton S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from <https://files.eric.ed.gov/fulltext/ED515656.pdf>
- Wise, A. E., & Darling-Hammond, L. (1984-1985). Teacher evaluation and teacher professionalism. *Educational Leadership*, 45(4), 28–33.
- Wise, A. E., Darling-Hammond, L., McLaughlin M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of the practices*. Prepared for the National Institute of Education. Santa Monica, CA: RAND.
- Witziers, B., Bosker, R. J. & Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Education Administration Quarterly* 39(3), 398-425.
- Wright, W. E. & Li, X. (2008). High-stakes math tests: How *No Child Left Behind* leaves newcomer English language learners behind. *Language policy*, 7(3), 237-266.
- Xu, X., Grant, L. W., & Ward, T. J. (2016). Validation of a statewide teacher evaluation system. National Association of Secondary School Principals. *NASSP Bulletin*, 100(4), 203–222.
- Yell, M. L., Katsiyannis, A., Collins, C. C. & Losinski, M. (2012). Exit exams, high-stakes testing, and students with disabilities: A persistent challenge. *Intervention in School and Clinic* 48(1), 60-64.
- Yen, W. M. (2007). Vertical scaling and No Child Left Behind. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 273–283). New York, NY: Springer.
- Zepeda, S. J., & Mayers, R. S. (2014). *Supervision across the content areas*. Retrieved from <http://ebookcentral.proquest.com>