THE RELATIONSHIP BETWEEN FLORIDA ACCOUNTABILITY PROGRAMS AND MERIT-BASED PAY IN TWO LARGE URBAN BASED SCHOOL DISTRICTS

By

Nicole Michelle Walkinshaw

A Dissertation Submitted to the Faculty of

The College of Education

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Florida Atlantic University

Boca Raton, FL

May 2022

THE RELATIONSHIP BETWEEN FLORIDA ACCOUNTABILITY PROGRAMS

AND MERIT-BASED PAY IN TWO LARGE URBAN BASED SCHOOL DISTRICTS

by

Nicole Michelle Walkinshaw

This dissertation was prepared under the direction of the candidate's dissertation advisor, Dr. Meredith Mountford, Department of Educational Leadership, and has been approved by the members of her supervisory committee. It was submitted to the faculty of the College of Education and was accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

SUPERVISORY COMMITTEE:

Meredith Mountford (Apr 4, 2022 16:40 EDT)

Meredith Mountford, Ph.D.
Dissertation Advisor

Daniel Reyes-Guerra, Ph.D.

John D. Morris (Apr 7, 2022 12:32 EDT)

John Morris, Ph.D.

Robert Shockley (Apr 7, 2022 12:36 EDT)

Robert Shockley, Ph.D.
Chair, Department of Educational Leadership
Research Methodology

Stephen Silverman (Apr 7, 2022 13:26 EDT)

Stephen Silverman, Ed.D.
Dean, College of Education

Robert W. Stackman Jr., Ph.D.
Dean, Graduate College

April 8, 2022
Date

VITAE

Nicole Michelle Walkinshaw is a career educator, activist, and patron of the Arts. She received her Bachelor of Science in Communication Processes and Disorders and a Bachelor of Arts for an Interdisciplinary Major in Women's Studies from the University of Florida. Later, she became a Fulbright Award Recipient to Japan. Nicole went on to obtain her Master of Education (M.Ed.) from Florida Atlantic University in Social Foundations in Multicultural Education with a concentration in ESOL (English for Speakers of Other Languages). She is also a National Board Certified Teacher (NBCT) in Adolescent Young Adulthood/English Language Arts (AYA/ELA). Nicole created multiple academic and artistic course programs that introduced many young adults to the realm of the Arts focusing on such notables as Rodin's The Thinker, Frida Kahlo, M.C. Escher, Langston Hughes, Charles Dickens, J.D. Salinger, Maya Angelou, Hayao Miyazaki, John Williams, Nina Simone, Ray Charles, and the Greats included in the acknowledgements. She was a NEA Florida Delegate to the National Convention & Representative Assembly while also maintaining ten Honor Society affiliations. In terms of Educational Leadership, Nicole went on to become a UCEA Jackson Scholar and will graduate Summa Cum Laude with a 4.0 GPA in her Ph.D. program. Her philosophical ideology is embedded in the notion that one must always fight for social justice so that malfeasance and ignorance will never become the norm.

"I am no longer accepting the things I cannot change.

I am changing the things I cannot accept." – Angela Davis

ACKNOWLEDGEMENTS

One of the most difficult things to do when putting pen to paper is to thank all of those who have been on this journey and have served in their own way as inspiration in one's work or in life. Not one element of this *process* could have been accomplished without the Divine helping to guide me and at times to carry me through pain, illness, doubt, and fear. I know that I am blessed to be here on this Earth and I hope with all that I am to show those who have believed in me that their time and efforts were not in vain and have served to mold the transformative educational leader that I have become through this rite of passage.

My family…My Mother – the Warrior Lioness who always protects her cubs and created strong women to maintain a legacy. My Father – the Intellectual Guiding Light who ensured that his children would learn to love reading, music, and the Arts. My Prodigy Sisters - Karen (the Mathematical Wizard who possesses magical powers) and Natalie (the Legal Phenomenon who understands the Law and Politics at an unparalleled level) have been and will always be the support pillars that have taught me about justice, equity, advocacy, artistic beauty, and meritocracy. They established a foundation of security and love that emboldened me to question the status quo, love the *process* of learning, and embrace the freedom to always do my due diligence in every encounter.

When one strives to create his or her own extended family, it is a wonder to genuinely comprehend those that are placed in your path. Dr. Meredith Mountford has been a guiding force who embodies authentic leadership, bravery, and has encouraged and

expected Excellence from her students at all levels. In the words of Descartes (I Think Therefore I Am) to Yoda (Pass On What You Have Learned), I will be forever grateful that she agreed to supervise my committee in order for me to truly become a forthright proponent and advocate for genuine transformative educational leadership and for advancements in educational policy. Dr. Daniel Reyes-Guerra, who initially introduced, informed, and encouraged me in this program, Dr. John Morris, a Master Mathematician, and Dr. Ira Bogotch, the first professor in my program, who taught me about the importance of context and academic writing while always maintaining a sense of Zen. All have been extraordinary with the support system they provided while sharing their insights regarding the impact Transformative Educational Leadership has upon the modern world.

For my students, colleagues, and friends, I want to give thanks as well. From the paideia seminars in class, the collegial discussions in the lunch room, or conversing about what it means to be an educator all served as sources of enlightenment.

In order to have reached this point, it has taken a village of many inspirational figures and while I may not be able to name them all, there are a few that I would like to acknowledge: Dr. Constance L. Shehan, Dr. Victoria Chen, Dr. Sylvie Blum-Reid, Dr. Leo Buscaglia, Dr. Carlos Diaz, Dr. Pat Maslin-Ostroski, Dr. John Pisapia, Dr. Grace Wolff, Dr. C. Everett Koop, Dr. Maya Angelou, Dr. Jacques Morcos, Dr. Jason Haffizulla, Dr. Zoyla Almeida, Dr. Fred Telischi, Dr. Niranjan Shintre, Dr. Charan Donkor, Dr. Rachel Maddow, Heather Cousino, Rouse Melvin Spencer, Iola Spencer, Irene White aka Dada, Ophelia Allen aka Mom, William Allen, Michael Walkinshaw, Nerissa Jemmotte, Liz Conforti, M.G. Angel, Von Rusty,

Joan Chan, Lyndon Robinson, Mary Foster, Paula Levendorf, Sherrel Brockington, Randi Weingarten, Melissa Slattery, and Dr. Joseph Morrissey. From the programs that I created, I want to thank Alex Haley, Kathy Bates, John Travolta, Terrence Howard, Jackie Chan, Jeffrey Katzenberg, John Lasseter, Jerry Bruckheimer, Rob Stewart, and the Cast & Crew of *Wicked*, *Lion King*, and *Hamilton*. May you all know my gratitude and love for encouraging, motivating, and sharing your knowledge and talent. It inspires me to do better each day.

In closing, Toni Morrison's words echo through my mind and must be shared:

> This is precisely the time when *Artists* go to work.
> There is no time to despair, No place for self-pity,
> No need for silence, No room for fear. We speak,
> We write, We do language. This is how civilizations
> Heal.

ABSTRACT

Author:	Nicole Michelle Walkinshaw
Title:	The Relationship Between Florida Accountability Programs and Merit-Based Pay in Two Large Urban Based School Districts
Institution:	Florida Atlantic University
Dissertation Advisor:	Dr. Meredith Mountford
Degree:	Doctor of Philosophy
Year:	2022

The purpose of this study was to ascertain insightful knowledge through the analysis of teacher pay across two similar K-12 public school districts that reflect the current evaluation methodology being utilized within the State of Florida. The two districts were selected because they are among the largest public school districts in the nation (Florida Department of Education [FLDOE], 2021a), have comparable student demographics and utilize contrasting weighted merit pay salary schedule profiles and algorithms which could provide insights into the relationship between accountability and merit-based pay. The Florida Standards Assessment (FSA) student assessments in (English [R%H] and Mathematics [M%H]) as well as the District School Grades (DSGs) were analyzed. The DSGs were commensurate with an approximation to the Value-Added Model (VAM) and Learning Growth Model (LGM) scores. Since the FLDOE does not release individual K-12 public school teacher VAM and LGM scores, the DSGs were the most appropriate comparative

score to utilize when comparing these two districts. These are the primary variables utilized by the Florida K-12 Public School Accountability Programs that directly impact merit pay salary schedule placement.

Quantitative methods employed statistical tests and analyses that included Independent Samples t-tests, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts between correlations that were calculated in the two K-12 public school districts. The study found that the input variables (R%H, M%H, and DSGs) had no statistically significant differences of the means between districts tested. Each input variable was commensurate over the 7-year study. Yet, there were statistically significant differences of the percentage mean in the output variables in regard to the percentage of K-12 public school teachers rated Highly Effective and Effective between both districts. The study also determined that the remaining input variable of the teacher classroom observation Instructional Practice (IP) score was significantly related to an educator's placement on annual merit pay-for-performance salary schedules. Both districts utilized the assessment tool iObservation ®. The results of all of the statistical analyses served to call into question the accuracy, legitimacy, and the efficacy of the methodology utilized to incentivize, compensate, and produce more Highly Effective educators.

DEDICATION

"*The paradox of education is precisely this – that as one begins to become conscious, one begins to examine the society in which he or she is being educated. Not everything that is faced can be changed but nothing can be changed until it is faced".*

~ James Baldwin

I dedicate this work to the Divine who imbues the soul with inspiration, creativity, and imagination. ~ "Believe you can and you're half way there". - Theodore Roosevelt.

I dedicate this work to my Father who is a champion of intellectual development, analytical fortitude, and academic achievement. ~ "Intelligence plus character, that is the goal of true education". – Martin Luther King, Jr.

I dedicate this work to my Mother who is the heartbeat and pulse that created daughters that would someday create Excellence. ~ "I have learned that people will forget what you said, people will forget what you did, but people will never forget how you made them *feel".* – Maya Angelou.

I dedicate this work to my sister, Karen, who instilled the importance of possessing a rational and analytical mind with a hint of whimsy. ~ "The ones who are crazy enough to think that they can change the world are the ones who do". – Steve Jobs.

I dedicate this work to my youngest sister, Natalie, who taught me the importance of the Law and the impact of activism in the political realm. ~ "What counts in life is not the mere fact that we have lived. It is what difference we have made to the lives of others that will determine the significance of the life we lead". – Nelson Mandela.

I dedicate this work to Dr. Meredith Mountford who believed in my ability and informed all of her students that we had to transform what it genuinely means to become a proactive Transformative Educational Leader. ~ “There are risks and costs to action. But they are far less than the long range risks of comfortable inaction”. – John F. Kennedy.

I also dedicate this work to Barbara Jordan who formed robust components of my educational ideology by stating that “the majority of the American people still believe that every single individual in this country is entitled to just as much respect and just as much dignity, as every other individual”.

Finally, I dedicate this work to Margaret Walker as she spoke to me during very challenging times. “For my people standing staring trying to fashion a better way from confusion, from hypocrisy and misunderstanding, trying to fashion a world that will hold all the people, all the faces and their countless generations; Let a new Earth rise. Let another world be born”.

THE RELATIONSHIP BETWEEN FLORIDA ACCOUNTABILITY PROGRAMS AND MERIT-BASED PAY IN TWO LARGE URBAN BASED SCHOOL DISTRICTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

The concept of advocating, engaging, and empowering individuals as coparticipants in decisions that directly affect their lives while acting with value and respect are the pillars that serve to define social justice (Blustein et al., 2001). Each construct must be embedded into the foundation of a healthy, sustainable, and productive educational environment. However, those who control the purse strings of power decided to systematically disenfranchise an entire class of professionals who are entrusted to mold our children into capable, confident young adults (Childress, 2019). According to Ritter and Jensen (2010), "performance pay does not capture all that teachers do" (p. 32). The National Center for Education Statistics [NCES] (2020) asserts that over the last two decades, Florida teacher pay has decreased 12.5% when adjusted for inflation. A report from the Economic Policy Institute produced by Garcia and Weiss (2019) confirms that teacher turnover is high largely because salaries are so low and pay raises so paltry. Moreover, the National Education Association (2021) maintains that "the average pay for teachers in Florida decreased, ranking the state 47th in 2019 to 49th in 2020" (p. 25). When viewed holistically, the contention could be made that this existential crisis will impede maintaining the framework of our democracy.

Ivashevskii (2011) further states, "the experience of history shows that ideological uncertainty leads to social crises" (p. 42). There have been considerable paradigm shifts in our American Educational System. A report from the Organization for Economic Cooperation and Development [OECD] (2017) which provides data on the structure,

finances, and performance of education systems throughout the world found that out of all of the OECD countries, the United States (U.S.) had the largest salary gap between teachers and other workers with comparable training. The gap between educator salaries and those with corresponding education in the U.S., referred to as "the teacher wage penalty, has grown since the 1990s from 1.8% in 1994 to a record 18.7% in 2017 when teachers made less than their counterparts" (Allegretto & Mishel, 2018, p. 3).

By broadening our contextual lens to focus upon our urban educational history, this research study endeavored to ascertain a more *contextualized* understanding of the impact of merit pay being linked to teacher performance. According to Kantor (2001), Tyack's *One Best System* offered a comprehensive history of the organizational revolution that transformed urban education in the United States between 1870 and 1940. The "*One Best System"* emerged from the organizational model of the factory and the large industrial corporation of modern America. Its highest values were "efficiency, rationality, continuity, precision, and impartiality" (p. 319). Tyack called this an "urban discipline" (Tyack, 1974, p. 29). The system was designed to replace village forms in which laymen participated in decentralized decision making with the new bureaucratic model of a closed 'non-political' system in which directives flowed from the top down, reports emanated from the bottom, and each step of the educational process was carefully prescribed by professional educators (Tyack, 1974). This model did not really value diversity or the dissenting opinions of others. It elevated the importance of objective testing by proffering the objectivity of the testing tool itself. Tyack (1974) emphasized the integral importance of sharing the decision-making power. He showed how values and structure are interconnected.

The contribution of teachers to student educational achievement has been demonstrated repeatedly by researchers (Rivkin et al., 2005; Rockoff, 2004; Kane & Staiger, 2008; Chetty et al., 2014). Subsequently, school districts and states increasingly have used teacher incentive pay programs as a tool for improving student performance. The motivation for these programs is to provide monetary incentives for teachers to increase effort that lead to higher measured student achievement (Brehm et al., 2017). However, impactful teaching faculty are discarded in the same way we discard vetted journalists who protect first amendment safeguards: by relabeling the work "content" and its workers "content providers" (Childress, 2019).

As of March 24, 2011, in conjunction with Florida's successful U.S. Department of Education (DOE)'s Race to the Top application, the former governor, Rick Scott, signed Senate Bill (SB) 736, sometimes referred to as the Student Success Act into law. This legislation required that all school districts in the State of Florida implement merit-based pay by June 1, 2014. Merit-based pay, sometimes referred to as pay-for-performance, was defined as any policy-based plans incorporated into a teacher's contract as a means of calculating a teacher's pay based on a demonstration of competence in teaching (DeSander, 2000) with 50 percent of their annual evaluation based on the Florida Comprehensive Assessment Test (FCAT) at that time.

Concurrently, SB 736 required that teachers hired during or after the 2011-2012 school year sign an annual contract. Veteran teachers, regardless of whether or not they earned multi-year contract security in the past through *tenure* or earning *permanent status* were now required to choose between their *tenure/permanent status* or jeopardize any future pay raises (Vagi, 2014). At the end of 2013, in Florida the FCAT test phased out in

lieu of the Post-Secondary Education Readiness Test (PERT) which was used to assess secondary student performance. However, in the 2014-2015 academic school year, the State Board of Education in Florida voted to eliminate the PERT test and implement the Florida Standards Assessment (FSA) exam. The FSA exam impacts primary through secondary students (3^{rd} – 11^{th}) who are on track for graduation. At present as well as through the end of the 2022 academic school year, the FSA exams are considered as a fundamental requirement for graduation and a high school diploma (Hollenbeck, 2018).

On September 14, 2021, Florida Governor Ron DeSantis, announced a legislative proposal that will eliminate the common-core based, end-of-year, high-stakes FSA and create the new Florida Assessment of Student Thinking (FAST) Test, which will monitor student progress and foster individual growth through three tests that will occur throughout the school year. By creating the FAST Test, Florida will become the first State in the nation to fully implement progress monitoring instead of end-of-year standardized testing and fully eliminate common core (Florida Department of Education [FLDOE], 2021d). With the implementation of this new testing paradigm, it does not actually end high-stakes testing and testing in the Spring term will still occur. Meier (2015) asserts that "Progress Monitoring Outcome Assessment (PMOA) possesses very nuanced distinctions. Regular monitoring of progress is a different procedure than examining outcomes" (p. 1). According to the Foundation for Florida's Future (2021) which was founded by former Florida Governor, Jeb Bush, a statement was released citing major concerns regarding these changes:

1. Does changing the nature of teacher-driven progress-monitoring tools create high-stakes stressors on students three times a year?

2. Will educators be required to teach on a schedule set by Tallahassee to be "on track" for three statewide progress monitoring tests?
3. Will the Spring progress monitoring test simply be a replacement for the end-of-year test and result in teachers having less time to cover the full year of content? (p. 2)

Moreover, according to Springer (2009), the use of the value-added merit pay systems in states like New York, Colorado, Tennessee, and Florida is becoming the standard. Value-Added Model assessments (VAMs) as well as alternative Local Growth Models (LGMs) both measure the effectiveness of a teacher based on students' academic growth from year to year primarily in Mathematics and English on the FSA test. In these instances, teachers' salaries are determined in part by their students' performance on standardized tests. Mayo (2012) states that in Florida, "the VAM formula makes Einstein's Theory of Relativity look like child's play" (p. 1). A major problem with such an ideological dispute is that it is often made in the absence of evidence of how such measures impact teacher performance and thus, affect student learning (Ritter & Jensen, 2010).

Statement of the Problem

It is important to question whether effective assessment practices persist in a performance pay regime. The fact that a principal identifies a teacher as "inadequate" on an evaluation observation does not necessarily mean that he or she will do so in a high-stakes environment. Undeniably, "a primary reason the single salary schedule replaced the grade-based compensation system was that subjective measures used to reward teachers were highly susceptible to gender and racial discrimination as well as nepotism"

(Marsden et al., 2007, p. 17). A major argument against merit-based pay programs concerns the difficulty in monitoring teacher performance. This research study endeavored to uncover new knowledge regarding the evaluation system for K-12 public school teachers in Florida and reveal the problems embedded within the state's triumvirate for evaluating K-12 public school teachers (i.e. Student Assessments - FSA English in the domain of Reading Percentage Satisfactory or Higher [R%H] and Mathematics Percentage Satisfactory or Higher [M%H], VAM/LGM scores which are commensurate with DSGs and Instructional Practice (IP) – Classroom Observation Scores). Accurate Observation of Teacher Performance (OTP) ratings "reflect a teacher's true effectiveness rather than any idiosyncrasies in judgment (i.e. biases and other rating errors) and lack of training or expertise applying the observation protocol" (Bergin et al., 2017, p. 19). Inaccurate ratings are unjust to teachers and provide misinformation on teachers' effectiveness globally. Additionally, it misidentifies particular strengths and areas needing growth, thereby failing both purposes of teacher evaluations. Capricious ratings are ethically unacceptable for high-stakes testing and personnel decisions (American Educational Research Association [AERA], American Psychological Association [APA] and the National Council on Measurement in Education [NCME], 2014).

The overall K-12 teacher rating relies upon empirical data that can be inconsistent, may be flawed due to subjectivity of the observation scores and is reliant upon a private equity firm to generate the student assessments currently and in the future. The Cambium Learning Group, Inc. creates all of the K-12 student assessment tests in Florida (i.e. FSA and FAST Progress Monitoring Tests) (FLDOE, 2021f) and that

company is a subsidiary which is owned by Veritas Capital, a multi-billion dollar private equity firm (Business Wire, 2018). However, this entity is not legally required to be transparent and maintain mandated oversight protocols that involve the development and dissemination of the FSA Test, the "creation of the new baseline data for accountability for the FAST Test" (National Center for Fair and Open Testing, 2021, p. 1), the implementation of the aforementioned student assessments, and its evaluation practices. According to the Public Education Director of the National Center for Fair and Open Testing (FairTest), Robert Schaeffer (2021), "the devil will be in the details on whether the move to progress monitoring will change the testing environment in Florida or if it is all just political posturing" (p. 2). In order to address the problems this study is trying to analyze, two large urban-based Florida K-12 public school districts were examined. They serve as a reflection and are indicative of the teacher evaluation methodology that is being utilized in K-12 public schools throughout the entire state.

Within the State of Florida, K-12 public school teachers are all placed upon merit pay-for-performance plans. Base pay salary movement does not occur unless a teacher receives an evaluation rating in either the Teacher Highly Effective Rating (THER) or the Teacher Effective Rating (TER) categories. In all K-12 public school districts within the state, teaching instructional personnel are evaluated in three domains. The domains include: IP, FSA Student Assessments (English in the domain of Reading % Satisfactory or Higher [R%H] and Mathematics % Satisfactory or Higher [M%H]) and VAM/ LGM scores – which are commensurate with an approximation to DSGs (FLDOE, 2020b). The Florida Personnel Evaluation Procedures and Criteria Act (2020) can be found in the Florida Statute §1012.34, subsections d and e, which identify those employed in the

instructional teaching fields that must have procedures and criteria for evaluation (FLDOE, 2020c). The state's platform premise that differentiates among the levels of performance is that if K-12 public school teachers display proficiency and exceptional performance, they will be rated THER as they would have accomplished the primary goal of attaining substantial and vigorous student achievement (FLDOE, 2020b).

The analysis of this underlying hypothesis was rooted in the presumption that if the empirical data that is reported to FLDOE for the Florida Accountability Programs (Florida School Recognition Program [FSRP] and Florida Differentiated Accountability Program [FDAP]) input variables of FSA (English in the domain of Reading [R%H] and Mathematics [M%H]) and the DSGs which are commensurate with an approximation to the VAM/LGM scores in both K-12 public school districts over a 7-year period, then the output variable data (THER and TER) should also be consistently comparable. Gayles (2007) asserts that the FSRP was created by the Florida legislature in 1997 to provide an incentive to schools that achieve specific standards-based goals. "Achievement is measured primarily through the state's high-stakes tests" (p. 439). By framing public schooling outcomes in terms of the "productivity" to be found in the "private sector" (FLDOE, 2004), it is problematic because it reproduces existing stratification. The more ominous implication that this policy makes is informed primarily by the work on reproductive aspects of schooling offered by Bourdieu and Passeron (1990) in addition to the relationship between the impact of the free-market model of schooling offered by the following researchers (Apple, 2018, 2001; Gerwirtz et al., 1995; & Bowles & Gintis, 2011; Lauder & Hughes, 1999; Whitty, 1997).

According to Murnane and Cohen (1986), “teacher performance is more difficult to monitor than performance in many other professions because output is not readily measured in a reliable, valid, and fair manner” (p. 9). Podgursky and Springer (2007) stated, “unlike, the sales of a salesman or the billable hours of a doctor or lawyer, the output of a teacher is not marketed” (p. 909). Thus, it is argued that the education sector cannot readily measure the value of the services provided by an individual teacher or group of teachers since achievement is influenced by many factors beyond the instructor’s control.

Another issue that was addressed focused upon how can teacher-leaders move up within this system with the barriers in place utilizing this type of observation methodology? According to Green and Oluwole (2015), “lawsuits had been filed against the state claiming that the teacher evaluation system is unfair because it partly rates their job performance on test scores of students they do not know and subjects they do not teach” (p. 402). Since 2009, thirty-six states and the District of Columbia have required teachers to be evaluated in part based on student scores on standardized tests. The idea has received a boost because of the Obama administration policies, particularly Race to the Top (Hazi, 2017). This represented a shift away from the federal control under the No Child Left Behind Act (U.S. DOE, 2011) and the Race to the Top (RTTT) that was passed as part of the American Recovery and Reinvestment Act (Committee on Health, Education, Labor, and Pensions, 2010), effectively requiring the use of VAMs and LGMs respectively.

According to Harrison and Cohen-Vogel (2012), “Senate Bills 6 and, later, SB 736, which became law in 2011, put forward ideas for substantively altering the teaching

profession in the State of Florida" (p. 521). Under the new law, teachers, traditionally compensated through pay scales that reward advanced education or experience, faced a new system in which performance-based evaluations drove salary increases. These new systems were required to base at least 50% of a teacher's evaluation on their value added to student achievement (Florida Senate, 2011). In addition, Senate Bill (SB) 736 eliminated the state's previous policy, which required that districts award "tenure" to teachers after 3 to 5 years of service by offering them long-term professional service contracts (Florida Contracts with Instructional Staff and Supervisors Act, [2010] – Florida Statute §1012.33 €). Under the new law, all teachers hired after July 1, 2011, are offered annual contracts that expire at the conclusion of each year, regardless of their length of service. A study was done by Polikoff et al., (2011) in regard to the accuracy of modern teacher evaluations which found that "test-based evaluation scores have little to no link to other teacher quality measures, such as how well instruction matches standards and the content of assessments" (p. 972). According to the Florida Personnel Evaluation Procedures and Criteria Act (2020) which can be found in the Florida Statute §1012.34 in relation to personnel evaluation procedures and criteria, the following is stated by the Florida State Legislature:

EVALUATION SYSTEM APPROVAL AND REPORTING

> For the purpose of increasing student academic performance by improving the quality of instructional, administrative, and supervisory services in the public schools of the State, the district school superintendent shall establish procedures for evaluating the performance of duties and responsibilities of all instructional,

administrative, and supervisory personnel employed by the school district. (Florida State Legislature, 2020)

The Florida Personnel Evaluation Procedures and Criteria Act (2020) which can be found in Florida Statute § 1012.34 places the power of constructing the parameters of the evaluation procedures in the purview of each district's school superintendent. The Beta Teachers Union filed a lawsuit against the School Board of Beta County because of the Superintendent's decision to unilaterally move teachers who had up to fourteen years of experience to the Merit Pay-For-Performance Plan without allowing those who had more than the five years of experience to choose whether or not they wanted to be grandfathered on to the prior pay scale salary model (Beta Teachers Union, 2014). The Department of Education's Inspector General (IG) (1996) stated that financial programs within education are "the most vulnerable to fraud, waste, and abuse in part due to the multitude of entities that assist in administering these programs" (p. 8). The IG also added that oversight from external sources should be explored in order to ensure genuine transparency with the process. In 2016, the Beta Teachers Union and the School Board of Beta County agreed to a Memorandum of Understanding (MOU), Court Case Number No. 4D15-1910, that grandfathered in those that were on the step schedule and made everyone else transfer to the pay-for-performance schedule (Florida District Court of Appeals, 2016). This served to inform the Beta veteran teachers that if they left the school system for any reason, if they chose to return, they would automatically be permanently placed on the pay-for-performance schedule thereby gradually eradicating the grandfathered pay schedule. At that time, the Pi District Schools also grandfathered their instructional personnel, however, they allowed the instructional personnel to choose

whether they wished to leave the aforementioned pay schedule and transfer to the pay-for-performance plan (Pi District Schools Compensation Department, 2014). According to Schwartz and Fayer (2006), “the equality required by democracy can easily be experienced as thin in contrast with the robust forms of substantive equality associated with that of distributive justice” (p. 293). All stakeholders must have a voice in the process.

According to Hammad (2010) in the school environment, respecting the collective decision-making insights of all of the stakeholders does not necessarily occur. In the shared decision-making (SDM) structure, “each participant should have an equal say in the process regardless of their position in the hierarchy. Nevertheless, not all teachers are included in this process” (pp. 97-98). Halpin and Troyna (1995) assert that “it would be conducive to define organizational culture in relation to the school environment” (p. 303). Torrington and Weightman (1989) define organizational culture as follows:

> The characteristic spirit and belief of an organization, demonstrated, for example, in the norms and values that are generally held about how people should treat each other, the nature of working relationships that should be developed and attitudes to change. (p. 519)

This suggests that cultural factors influence significantly how an organization functions; schools are no exception. Like other organizations, schools are strongly influenced by the norms and values held by their members (Dalin et al., 1995). According to Fertig’s (2014) work, he contends that school culture shapes school management behaviors, which are important aspects of the school as an organization. “It is when a

teacher's pay is linked to their performance rating that lines begin to blur and the cultural environment is impacted" (Fertig, 2014, p. 199). Wragg et al., (2000) ponder whether:

> If those teachers who are stigmatized as weak actually do fail; and Do teachers see the problems differently from their superiors? If employees are paid for performance in the private sector, why should a similar approach to compensation not work for public school teachers? (p. 86)

Such reasoning apparently persuaded President Reagan to throw his support behind the drive to introduce merit pay in school systems (Duke, 2005). Performance evaluation and accountability have long been central values in educational management (Shinkfield & Stufflebeam, 1995).

Garms (1986) stated, "the art of teaching is such that valid, objective, measurable criteria cannot be found and implemented in teacher evaluations" (p. 371). In relation to this study, the FLDOE (2020b) specifies that "statewide student assessments are developed, administered, scored, and reported by two different assessment vendors" (p. 2). The American Institutes for Research (AIR) is the vendor that impacts the K-12 public school domain in that it administers the FSA, which are aligned to the Florida Standards in English Language Arts (ELA) and Mathematics (Florida Standards Assessment, 2015). According to AIR (2021), it was established in 1946, with headquarters in Washington, D.C., is a nonpartisan, not-for-profit organization that conducts behavioral and social science research both domestically and internationally in the area of education. However, it should be noted that AIR reached an agreement with the Cambium Learning Group, Inc. At the end of 2019, AIR agreed to sell off its student assessment division (AIR, 2021). Furthermore, according to Business Wire (2018) as of

December 2018, the Cambium Learning Group, Inc is now owned by Veritas Capital, a New York-based private equity firm. According to the multi-billion dollar Veritas Capital's (2020) Team Page, it states that it is comprised of 50 employees. Historically, private equity firms have had minimal regulatory oversight because their investors were mostly high-net-worth individuals (HNWI) - "who were better able to sustain losses in adverse situations and thus required less protection" (Bierman, 2011, p 3). Ironically, as of September 14, 2021, Florida Governor Ron DeSantis and the FLDOE have posted that it will be the Cambium Learning Group's Progress Monitoring Assessment that will create the three yearly tests that the students will have to take as of the 2023-2024 academic school year as a graduation requirement that is still linked to high-stakes testing outcomes (FLDOE, 2021d). Florida Education Commissioner, Richard Corcoran, has assured that the high stakes accountability measures, like school grades, teacher evaluations, "turnaround", third grade retention, etc., will remain but will be eventually attached to the scores achieved on these new state mandated progress monitoring assessments (FLDOE, 2021c). On February 15, 2021, J. Alex Kelly, Chief of Staff at the FLDOE, testified before the Florida Senate Select Committee on Pandemic Preparedness and Response. Mr. Kelly reported how the FLDOE spent federal Coronavirus Aid, Relief, and Economic Security (CARES) funding to provide data driven support for progress monitoring. The FLDOE spent $10 Million of CARES funding purchasing a progress monitoring tool, the Adaptive Progress Monitoring, APM, - developed by Cambium Assessment, Inc., the same contractor for the FSA, for districts to use "free of charge" (Florida Senate, 2021).

It becomes a question of how teachers can be held accountable and student performance linked to their pay when there are a myriad of other variables that can impact student performance like poverty, home environment, peers, as well as the instrument and vendor that are used for student assessment. There may be no singular viable solution to the issues concerning merit pay, however, this matter definitely necessitates further investigation.

The use of DSGs which are commensurate with an approximation to VAM and LGM scores that measure teacher effectiveness are based on several theoretical and methodological assumptions about measuring the contribution of a teacher to the learning of his or her individual students. Teacher evaluations have been central to the national reform agenda to improve teacher quality. NCLB (2002) required states to have highly qualified teachers and RTTP funds (U.S. Department of Education, 2016) required states to have highly effective ones and to improve their effectiveness through teacher evaluations (Pullin, 2013). By definition, VAMs and LGMs are designed to isolate and measure teachers' contributions to student learning and achievement on large-scaled standardized tests as groups of students move from one grade level to the next. Statisticians measure value-added by "mathematically calculating the "value" a teacher "adds to" or "detracts from" student achievement scores over time, and as compared to teachers with "similar" students" (Amrein-Beardsley et. al., 2013, p. 3). Purportedly, VAMs and LGMs allow for richer analyses of achievement data by tracking student learning trajectories from the time they enter a classroom to the time they leave. (Harris, 2011).

Primarily, the value-added estimates of teacher effects are treated as measures of "teacher effectiveness" (Berliner, 2014; Braun, 2005; Corcoran, 2009). From that perspective, it is deplorable to maintain the assumption that teachers who positively impact individual students also have the same effect on entire classes of students simplifies "the complex interactions of numerous in-and out-of-classroom/school exogenous variables to a presumably one-directional relationship between teachers and their students" (Berliner, 2014, p. 15). Additionally, Berliner (2014) maintains "there is growing evidence that innumerable and often invisible variables exist that confound the attribution of a student's test score to his or her teacher" (p. 22).

VAMs and LGMs claim to measure teacher effectiveness based on the assumption that a student's performance on a valid, reliable test measures his or her mastery of the aligned curriculum (Corcoran, 2009; Shavelson & Marsh, 1986; see also Little et al., 2009). The student's mastery is then attributed to teacher behaviors, which once again is assumed to be a presumably valid, reliable measure of the teacher's effectiveness (Shavelson & Marsh, 1986; see also Little et al., 2009). In order to evaluate a teacher evaluation system, it is essential and critical "to examine different types of validity evidence with regards to the use of tests to make inferences about teacher quality" (Herlihy & Corey, 2014, p. 11). In the State of Florida, Governor Ron DeSantis signed House Bill (HB) 641, Funds for the Operation of Schools, and announced the approval of $500 million in the State's budget that is dedicated to raising teacher salaries. $400 million is invested to raise the minimum base pay for full-time starting classroom teachers, and $100 million is to raise the salaries of Florida's veteran teachers and other instructional personnel (FLDOE, 2020c). The largest percentage of these monetary funds

are not for veteran teachers as the bulk of this budgetary capital is ladened towards newly hired teachers at the entry level (HB 641, 2020).

Background of the Problem

Different narratives have been constructed about the nature of the learning process and the causes of poor student achievement. These competing causal narratives are central to efforts to portray performance pay as a positive force for improving student outcomes (e.g., Cochran-Smith, 2001; Cochran-Smith & Fries, 2001; Cohen-Vogel, 2005; Cohen-Vogel & Herrington, 2005; Earley, 2000). In terms of this research, pay-for-performance merit-based pay salary schedules for K-12 public school teachers were explored as the State of Florida's method of reforming the K-12 teaching profession. In Florida, the two large urban school districts, Beta and Pi, were selected because they are ranked as two of the largest school districts in the nation (FLDOE, 2021a), have comparable student demographics and utilize contrasting weighted merit pay salary schedule profiles and algorithms which could provide insights into the relationship between accountability and merit-based pay. Through a quantitative lens, it is critical to pose the supposition that if there is no significant difference in the input variables: FSA R%H, M%H, and DSGs which are components of the Florida Accountability Programs that relate directly to merit pay, then there should be no significant difference in the output variables of the teacher evaluation ratings in the categories of THER and TER when analyzing two commensurate large urban K-12 public school districts. Phelps et al. (2005) assert that "quantitative methods greatly enhance the study and the data are in a better initial position to produce research that is significant to the field of education" (p. 95).

According to Şen et al. (2020), the VAM and LGM scores, which are a part of the teacher evaluation process, “use a complicated mathematical formula to estimate how much a specific teacher helped specific children improve on test scores beyond how that student was already expected to perform” (p. 79). Since the 1800s, there have been only three major changes in the method of teacher pay: “an initial rural tradition of paying teachers room and board, a move to a grade-based salary schedule, and the shift to today's single salary schedule including more recent compensation reform attempts, such as the merit pay plans and career ladder programs” (Protsik, 1996, p. 267). In the 1860s, job requirements instead “focused on basic knowledge of the 3Rs (reading, writing, and arithmetic), and possession of "certified moral character" and a middle-class appearance” (Tyack et al., 1981, p. 131).

In 1902, while some individual teacher bargaining with school board members still occurred, “it became more common for cities to establish salary schedules for teacher pay” (Tyack et al., 1981, pp. 142-143). “Many states adopted a minimum salary level below which no teacher could be paid, in an attempt to address the problem of high teacher turnover” (Frase, 1992, p. 5). Individual cities, meanwhile, set position-based salary schedules (also called differentiated salary schedules) for men and women, and for whites and blacks. According to English (1992), teachers were “paid based on their years of experience, gender, race, and the grade level that they taught” (p. 9). “School administrators could also factor a subjective measure of merit into the teachers’ salaries” (Tyack et al., 1981, p. 150).

The differentiated salary schedules of the early 1900s contained a merit component. Wilmers and Massenkoff (2020) assert that due to wage stagnation employers

experimented with practices, like merit-based pay increases, that seemed likely to raise pay. The declining real value of the minimum wage is often cited as a key determinant of stagnation of earnings distribution (Autor et al., 2016). Research has suggested that merit pay may tend to raise, not lower wages. Hence, merit-based pay is part of a set of effective management practices that drive increased productivity (Bloom & Van Reenen, 2007).

Yet the 1980s were the first period where there was a national call for improving teacher performance through monetary incentives. “President Reagan led the way, making merit pay one of his "bully pulpit" issues” (Frase, 1992, p. 5). Bonuses equal to nine percent of salaries were awarded each year for four years to teachers deemed “skillful” or “exemplary” (Hatry et al., 2012, p. 132). While there is public support for these plans, “the major teacher organizations believe these plans are too costly and that they undermine teacher collegiality by creating a competitive work environment” (Richardson, 1994, p. 45). These criticisms are not unlike those of other merit pay plans across the nation. According to a 1979 Educational Research Service Study, “most merit pay plans are discontinued within six years, largely due to problems of administration and personnel, collective bargaining, and budgetary shortfalls” (Murnane & Cohen, 1986, p. 5). One of the most prominent employment practices changes is “the spread of merit-based pay which is defined as variable changes in base pay (not one-time bonuses) that are determined through subjective assessments of performance practices by managers and administrators” (Wilmers & Massenkoff, 2020, p. 179).

Overall, in a study of eighteen school district programs since 1983, Hatry et al. (2012) found that “most school districts which implemented incentive plans for teachers

were unsuccessful in creating lasting and effective programs" (p. 133). Most districts cited significant teacher morale problems stemming from competition, unfair evaluation practices, and the use of quotas in determining the number of teachers to receive awards. "Programs were also costly (when funding was stable), and were difficult to administer" (Hatry & Greiner, 1984, p. 32). "The history of teacher pay demonstrates how compensation systems have changed over time to meet newly developed needs, such as addressing high teacher turnover rates, and meeting the women's movement call for equal pay" (Protsik, 1996, p. 271).

Data from national surveys done from as far back as 2006 show that "close to 100% of traditional K-12 public school teachers are employed in school districts that make use of salary schedules in pay setting" (Podgursky & Springer, 2007, p. 909). Therefore, "roughly 3.1 million public school teachers from kindergarten through the secondary level are paid largely on the basis of years of experience and education level—two variables weakly correlated, at best, with student outcomes" (Hanushek, 2002, p. 65). Although merit-based pay programs date back to Great Britain in the early 1700s, and "somewhat similar ideas formed around the notion of performance contracting in the late 1960s" (Stucker & Hall, 1971, p. 8), it was not until the release of the *A Nation at Risk* report in 1983 that a significant number of public school districts in the United States began considering merit-based pay as an alternative or supplement salary schedule. Merit-based pay rewards individual teachers, groups of teachers, or schools on any number of factors, including, but not limited to student performance, classroom observations, and teacher portfolios. "Merit-based pay is a reward system that hinges on

student outcomes attributed to a particular teacher or group of teachers rather than on "inputs" such as skills or knowledge" (Podgursky & Springer, 2007, p. 909).

Purpose of the Study

The purpose of this study was to ascertain insightful knowledge through the analysis of teacher pay across two similar K-12 public school districts that reflect the current teacher evaluation methodology being utilized within the State of Florida. The selected districts served as substantial indicators of the teacher evaluation methodology implemented in K-12 public education throughout the entire state. This research study sought to determine if the districts maintained comparable student assessment scores in R%H, M%H, and DSGs which represent commensurate VAM and LGM scores. If there were no significant differences of the means in each of the input variable categories between districts, then no disparity should exist in the teacher rating outcome scores of THER and TER that determine placement upon K-12 pay-for-performance salary schedules. These two districts possessed equivalent student demographics, should produce comparable FSA results in the pivotal R%H and M%H categories and should have comparable DSGs. DSGs are commensurate with an approximation to the VAM and LGM scores. Since the FLDOE does not release individual K-12 public school teacher VAM and LGM scores, the DSGs were the most appropriate comparative score to utilize when comparing these two districts. These are also the primary variables utilized by the Florida K-12 Public School Accountability Programs (FSRP and the FDAP) that directly impact merit pay.

Quantitatively this means that this study analyzed if there was a statistically significant difference of the means between districts in the (x)-input variable domains of

FSA data in R%H, M%H and DSGs. These are components of the Florida Accountability Programs: the FSRP and the FDAP that fund and impact teacher merit pay salary schedules. Moreover, this research investigated if any percentage differences of the means between districts existed in the (*y*)-output variables of teacher instructional personnel evaluation ratings in the THER and the TER category domains which currently serve as predictors of authentic student achievement in relation to merit pay incentivization. These are the only two rating categorical domains that qualify for placement out of the five-tier evaluation system for K-12 public school merit pay plans in the state.

Moreover, the FLDOE and the districts do not publish individual teacher VAM/LGM scores. For this reason, DSGs were analyzed as they are commensurate with an approximation to individual teacher VAM/LGM scores. The input and output variables were examined and analyzed between two comparable large urban based K-12 public school districts in Florida.

Research Questions

There were five empirical research questions and five null hypotheses that were examined throughout this study to determine the validity of linking the outcome variables of teacher merit-based pay-for-performance (THER and TER) to the input variables (FSA R%H, M%H, and DSGs) that are components of the Florida Accountability Programs (FSRP and the FDAP). The researcher also endeavored to determine if extreme deviations existed in the personnel evaluation output variables (i.e. [THER] and [TER] Rankings) between the two comparable large urban based Florida K-12 Public School Districts that would severely impact their annual salary.

Therefore, this research posited the following empirical research questions:

1. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA English Domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts?
2. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA Domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts?
3. Is there a statistically significant difference of the means between the (*x*)-input variable of the District School Grades (DSGs) between two large urban based Florida K-12 Public School Districts?
4. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (*y*)-output variable of the Highly Effective (THER) Category Rating Designation between two large urban based Florida K-12 Public School Districts?
5. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (y)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts?

Null Hypotheses

The five empirical research questions were addressable by a null hypothesis with the K-12 public schools from both districts serving as the unit of

analysis. The K-12 Instructional Staff Teacher Evaluation Rating data are reported to FLDOE. Therefore, this research posited five null hypotheses:

$H_0$1. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA English domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts.

$H_0$2. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts.

$H_0$3. There is no statistically significant difference of the means between the (*x*)-input variable of the DSGs between two large urban based Florida K-12 Public School Districts.

$H_0$4. There is no statistically significant difference of the means between the (*y*)-output variable of the Highly Effective (THER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

$H_0$5. There is no statistically significant difference of the means between the (*y*)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

With the impact of the current global pandemic of Coronavirus (Covid-19), it is unclear as to whether the scope and spectrum of this deeply entrenched model of pay-for-performance will be sustainable.

Conceptual Framework for Florida K-12 Teacher Evaluations

The following section of the paper discusses the evaluation systems utilized in each K-12 public school district analyzed. It provides a foundation for how teachers are classified for salary schedule placement through the merit pay plans.

Beta District - Instructional Development and Growth Evaluation System

According to the FLDOE in 2021, the Beta Instructional Development and Growth Evaluations System (BrIDGES) for Beta County Public School District (BCPS) is designed to foster high quality instruction and increase student achievement. This plan outlines how this system has been used from the 2012-2022 academic school years. Evaluation data is collected and analyzed in order to make decisions to determine whether this increases teacher effectiveness and impacts student achievement. This is monitored through the use of IP Scores (i.e. Highly Effective, Effective, Needs Improvement, Developing and Unsatisfactory Rating Designations), generated by individual element ratings through the Marzano Observation Tool iObservation ®, Student Assessment Performance Scores and District VAM/LGM scores which are represented by a commensurate approximation to DSGs (FLDOE, 2020a).

Pi District - Teacher Evaluation Model

The Pi District Schools also use the Marzano Focused Teacher Evaluation Model iObservation ® to generate corresponding scores (Pi District Schools Instructional Evaluation System, 2018). In both K-12 public school districts, the iObservation ® model

serves to process the data input which in turn impacts how the teachers are rated Highly Effective, Effective, Needs Improvement/Developing, Unsatisfactory. It is on the basis of the distribution of the two initial ratings that are provided by the districts to the FLDOE from the reported instructional personnel evaluations in conjunction with VAM/LGM scores represented by the commensurate approximation of DSGs and FSA student assessments that determine placement upon the merit pay-for-performance teacher instructional salary schedules (FLDOE, 2020c).

Marzano Classroom Focused Teacher Evaluation Model

Schön (1987) found that effective teachers were reflective about their practice. He described "reflection-in-action" as a process where tacit knowledge is replaced by a "surprise" response. According to Baker and Rozendal (2019), in education, teachers must engage in reflective practice to address the ever-changing and surprising responses that students have toward their instruction. Effective teaching goes beyond implementing a district curriculum. "Teachers must learn to determine students' learning needs, identify and implement activities and methods that will meet these needs, and reflect on the effectiveness of their implementations" (p. 59). Marzano and Kendall (2007) argued that "Bloom's taxonomy was based on a faulty notion of hierarchical cognition" (p. 46). Additionally, Shulman (1998) contended that reflecting on practice and making changes to actions improves teacher and student productivity. "In education, teachers engage in reflective practice to address the ever-changing and surprising responses that students have toward their instruction" (p. 7). Reflection as a learning process has been conceptualized as the cyclical process of asking oneself (a) what, (b) so what, and (c) what's next? (Rolfe et al., 2001). Marzano and Kendall (2007) posit that even though

Bloom's taxonomy was used to formulate learning objectives, design tests and evaluations, and develop curricula, "it is the metacognitive dimension that is critical because it includes self-knowledge which embodies beliefs, goals, and motivations that reside in the learner" (p. 61). Consequently, Marzano endeavored to "create a cognitive-based rubric for teachers to reflect and to formulate learning objectives that span the spectrum of learning" (Baker & Rozendal, 2019, p. 62). This evolved into the construction of a performance-based rubric whereby teachers and students were assessed based upon their performance (Anderson et al., 2014). However, with the passage of the American Recovery and Reinvestment Act of 2009 (Recovery Act), "the Race to the Top Funds allowed states to compete for funds in order to reward highly effective teachers" (Honawar & Olson, 2008, p. 26). Although a consensus exists on what effective teachers do to enhance student learning, there are also meta-analyses by researchers who have developed instruments (i.e. the iObservation ® tool) to quantify subjective qualitative input to determine the effects of specific instructional strategies (Marzano et al., 2008). While teaching undeniably will remain an art, there is also a science to it that is being aggressively applied to practice (Tucker & Stronge, 2005).

According to the Florida Senate (2011), statutes contained in SB 736, the Student Success Act, required that IP be evaluated for classroom teachers in direct relation to the Florida Educator Accomplished Practices (FEAPs). The key components of the input variable of IP were evaluated using the Marzano Teacher Evaluation Model (MTEM) (FLDOE, 2014).

Although pay-for-performance plans are becoming the standard throughout the United States (U.S.), they are not currently supported by research. To date, "there are no

rigorous empirical validations to show that U.S. performance-pay programs in education are linked to substantial and sustained successes, either in elevating student achievement or attracting a wider pool of able teacher candidates" (Wayne & Youngs, 2003, p. 91). Cissell (2010) asserts that although legislation has been passed that allows the tracking of individual students' progress on a performance measure and linking that performance to teachers, "research has yet to determine the authentic effectiveness of these teacher pay-for-performance programs across the nation" (p. 127).

The Florida Model Framework that utilizes the Marzano iObservation ® tool is used in the Beta School District (BSD, 2021b) for all K-12 public school classroom teachers at all school sites. This model has a total of 23 elements, aligned to the Florida Educator Accomplished Practices (FEAPs) adopted by the Florida State Board of Education.

> There will be ongoing observations during the school year with at least 1 observation cycle to collect growth and evaluation data. Once the ratings from the elements scored have been equally averaged, the IP Score will be generated, which is worth 50% of the overall evaluation. This IP score will be combined with Student Performance (SP) that is worth 35%, and Deliberate/Professional Practice (DP) that is worth 15% in order to obtain an overall evaluation score. (FLDOE, 2021a)

The Pi School District Model of Instructional Evaluation is also rooted in elements of Marzano. In 2011, the School District of Pi County and the Classroom Teachers Association staff met and reviewed the State's suggested research-based evaluation/observation systems. The District and the Classroom Teacher Association

entered into a MOU that created a joint negotiations committee to develop a new evaluation instrument in accordance with the new State Statute. The MTEM was selected, recommended to, and was ultimately approved by the School Board (FLDOE, 2021b).

In 2018, the School District of Pi County and the Classroom Teachers Association staff approved an update from the MTEM to the Marzano Focused Teacher Evaluation Model where it was adapted for the Pi District Schools to become the Pi Model of Instruction. (FLDOE, 2021a). In the Pi School District, the classroom teacher model is a process which includes performance indicators that focus on four domains. The four domains of the Pi Model of Instruction contain 22 total elements and build on each other to support teacher growth, development, and performance. Unlike other evaluation models, this is a coaching model that has been shown in causal studies to have the most direct effect on student performance.

> Together, the four domains contain 22 elements that define a knowledge base for teaching and a framework for the systematic development of expertise. The Evaluation System is made up of three scoring components, IP, Student Performance (SP) and Professional Practice (PP). Each component weighting the same at one-third (33.3%) (FLDOE, 2021a).

Methodology

This study was designed to analyze teacher pay across two comparable K-12 public school districts that served to reflect the evaluation methodology being utilized within the State of Florida. These two districts possessed equivalent student demographics, produced comparable FSA results in the critical R%H and M%H categories and have comparable DSGs. DSGs are commensurate with an approximation

to the VAM and LGM scores. Since the FLDOE does not release individual K-12 public school teacher VAM and LGM scores, the DSGs were the most appropriate comparative score to utilize when comparing these two districts. These are also the principle variables utilized by the Florida K-12 Public School Accountability Programs (FSRP and the FDAP) that directly impact merit pay. Quantitatively this means that this study analyzed if there was a statistically significant difference of the means between districts in the (x)- input variable domains of FSA data in R%H, M%H, and DSGs. These are components of the Florida Accountability Programs: the FSRP and the FDAP that fund and impact teacher merit pay salary schedules. This research also investigated if any percentage differences of the means between districts existed in the (y)-output variables of teacher instructional personnel evaluation ratings in the Highly Effective Rating and Effective Rating category domains which currently serve as predictors of authentic student achievement in relation to merit pay incentivization. These are the only two rating categorical domains that qualify for placement out of the five-tier evaluation system for K-12 public school merit pay plans in the state. Moreover, the FLDOE and the districts do not publish individual teacher VAM/LGM scores. For this reason, DSGs were analyzed as they are commensurate with an approximation to individual teacher VAM/LGM scores.

The study further examined whether the output variables of teaching instructional personnel evaluation ratings in the categories of THER and TER annually showed significant differences of the mean percentage between the K-12 public school districts of teachers rated in each respective category. Independent Samples t-tests were run for Means and Standard Deviations comparisons on the input variables. As an addendum,

Bivariate Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were constructed to ascertain "information regarding the difference in correlations and the relationships between the input variables" (Loesch, 1986, p. 288). The output variables were analyzed based upon the K-12 public school district data reported to FLDOE. The tests and analyses were implemented to meet the purpose of this study. More specifically longitudinal analyses of 7 years (2012-13; 2013-14; 2014-15; 2015-16; 2016-17; 2017-18; 2018-19) of empirical data occurred utilizing FLDOE data from the Beta and Pi School Districts in relation to the input variables of FSA R%H, M%H and DSGs. The output variables of teacher instructional personnel evaluation rating scores (5-tier System) were collected in the domains of Highly Effective (THER), Effective (TER), Needs Improvement (NI), 3-Years Developing (3-YD), and Unsatisfactory (U) (FLDOE, 2020c).

> Pursuant to the FLDOE Emergency Order No. 2020-EO-1, the Spring K-12 statewide assessment test administrations for the 2019-20 school year were canceled due to Covid-19 and accountability measures reliant on such data were not calculated for the 2019-20 school year. (FLDOE, 2020c)

Research Design and Methods of the Study

This section examined the FLDOE's databases that related to the output variables (THER and TER) through data analyses and Independent Samples t-tests using all of the K-12 public schools contained within both districts as the unit of analysis for the annual K-12 public school teacher instructional personnel performance rating evaluations. The input variables of FSA student assessment data (R%H and M%H) and DSGs which are commensurate with an approximation to VAM/LGM scores also utilized all of the K-12

public schools contained within both districts as the unit of analysis for Independent Samples t-tests. Additionally, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were constructed to ascertain any information regarding any relationships that may exist between the input and output variables in each individual district and between districts. The information was collected digitally and analyzed to determine any correlational relationships.

Assumptions

For the purposes of this quantitative analysis, the following assumptions were made:

1. The State's archival databases included K-12 public school data relating to teacher performance rating categories utilized for annual teacher evaluations.
2. The State's archival databases included K-12 public school data relating to FSA English and Mathematics derived from student assessments.
3. The State's archival databases included K-12 public school data relating to the reported DSGs.
4. The data for the K-12 public school districts were accurately recorded by the State.

Site Sample

The sample of this study included the annual district teacher instructional personnel evaluation score ratings by school (THER and TER) for K-12 public school teachers as well as the FSA student assessment (R%H and M%H) data and DSGs from

two comparable large urban K-12 public school districts in Florida (i.e. Beta and Pi). In these two districts the data reflected the K-12 public schools. All schools classified as a State of Florida K-12 public school was included in this data.

Data Collection

The FLDOE annually publishes K-12 public school teacher instructional personnel rating evaluations by school, FSA in English and Mathematics (R%H and M%H) as well as DSGs by district. The data is archived on the department's website and is readily accessible to the public. Thus, all of the pertinent data relating to Florida's Accountability Programs (FSRP and the FDAP) from these large urban K-12 public school districts were retrieved digitally.

Data Analysis

An Independent Samples t-test, a Means and Standard Deviation comparison, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts for all of the input variables (FSA R%H, M%H, and DSGs) and output variables (THER and TER) were performed on all of the K-12 public schools contained within each district annually (unit of analysis) were performed. The data analyses were performed on the input variables and the output variables of two comparable large urban K-12 public school districts that is reported to the FLDOE.

The input variable domains of R%H, M%H, DSGs and the output variable domains of Teacher Instructional Personnel Rating Evaluation Categories of THER and TER were analyzed to determine if either of the two comparable large urban K-12 public school districts' teacher evaluation rating category designations for merit pay salary schedule placement (i.e. THER and TER) exhibited significant statistical difference of

the means between districts over a seven-year period thereby revealing any relevant information in relation to pay-for-performance for the K-12 public school educators.

Limitations

1. Although the FLDOE archives the personnel evaluation information for the K-12 public school teachers, FSA English and Mathematics data as well as DSGs, it is not currently updated from the previous school year due to the Covid-19 pandemic thereby hindering the most up-to-date information that the researcher can acquire.
2. The sample in this study was limited to the State's K-12 public school district teacher instructional personnel evaluation ratings, K-12 public school FSA English and Mathematics data and K-12 public school DSGs. Private school information was not accessible for comparison in this study.
3. A small number of students with disabilities who have an Individual Education Plan (IEP) may not be required to take the FSA Student Assessments. Only students who meet the exclusion criteria set forth in the State Board Rule 6A-1.0943, Florida Administrative Code, Statewide Assessment for Students with Disabilities, can be excluded from taking the FSA Student Assessments. (FLDOE, 2021c)

Delimitations

1. This study specifically focused upon two large urban K-12 public school districts in the State of Florida which was limited to a seven-year analysis of the K-12 teacher instructional personnel evaluation score rating

domains, K-12 public school FSA English and Mathematics data and K-12 public school DSGs and its relationship to placement on merit-based pay salary schedules.

2. The FLDOE Emergency Order No. 2020-EO-1, the Spring K-12 statewide assessment test administrations for the 2019-20 school year were canceled due to the pandemic outbreak of Coronavirus and accountability measures reliant on such data were not calculated for the 2019-20 school year. (FLDOE, 2020c)

Definitions

For the purposes of this study, the five-tier levels of performance for the teacher evaluation category designations were defined as follows:

Highly Effective - The teacher continuously identifies and highlights the content that is critical for students and, by the end of the lesson, these efforts portray a clear progression of content that leads to deeper understanding of the content. The teacher also establishes appropriate student growth goal(s) for subgroups of students not reaching full potential in collaboration with students, parents, and other school staff. Goal(s) identify multiple, high-quality sources of data to monitor, adjust, and evaluate achievement of goal(s). (Marzano et al., 2013, p. 6)

Effective - The teacher identifies and effectively employs interventions that meet the needs of specific subpopulations (e.g., English Language Learners [ELL], special education, and students who come from environments that offer little support for learning). The teacher also establishes appropriate student growth goal(s) for subgroups of students not reaching full learning potential. Goal(s)

identify multiple, high-quality sources of data to monitor, adjust, and evaluate achievement of goals. (Marzano et al., 2013, p. 6)

Needs Improvement - The teacher identifies interventions that meet the needs of specific subpopulations (e.g., ELL, special education, and students who come from environments that offer little support for learning), but does not ensure that all identified students are adequately served by the interventions. Additionally, the teacher establishes appropriate student growth goal(s) for subgroups of students not reaching full learning potential. However, the goals do not identify multiple, high-quality sources of data to monitor, adjust, and evaluate achievement of goal(s). (Marzano et al., 2013, p. 6)

3 Years – Developing – Implemented by the State of Florida regarding the first three years of a new teacher's performance during his or her initial probationary period if the teacher does not meet the standards required of the Highly Effective or Effective teacher evaluation category rubric. (FLDOE, 2014, para. 3)

Unsatisfactory - The teacher does not know or understand the intervention system or does not use the intervention system to address student needs. The teacher also does not establish student growth goal(s) or establishes inappropriate goals for subgroups of students not reaching full learning potential. The goal(s) do not identify multiple, high-quality sources of data to monitor, adjust, and evaluate achievement of goal(s). (Marzano et al., 2013, p. 6)

Furthermore, as set forth in Rule 6A-1.09981, Florida Administrative Code (FAC), and the Florida Grading System Act (2021) can be found in the Florida Statute § 1008.34. Thereby, the DSGs for each of the counties in Florida were assigned letter

grades (i.e. A, B, C, D, F) based upon a numeric range. In order to determine a school grades calculation, the points earned for each component are added together and divided by the total number of available points to determine the percentage of points earned. The school grading percentages encompassed the following:

A = 62% of points or greater

B = 54% to 61% of points

C = 41% to 53% of points

D = 32% to 40% of points

F = 31% of points or less (FLDOE, 2021e)

In order for any K-12 public school to attain a grade, every school must test 95 percent of their student population (FLDOE, 2020b).

Significance of the Study

The significance of the study was to determine if there was consistency between districts regarding the input variables of the Florida Accountability Programs (FSA R%H, M%H, and DSGs) over a 7-year period. These input variables are the principle factors that impact student achievement, teacher evaluation rating designations, and merit pay. The hypotheses postulated that if there was consistency over the 7-year duration in each input variable category between districts, then would the overall teacher evaluation ratings be divergent between both comparable K-12 public school districts within the state. The IP score is supposed to be reflected in the results of the FSA (R%H and M%H) and the VAM/ LGM scores represented by the commensurate approximation to the DSGs scores. Santora (2019) believes that teachers become demoralized when they "cannot do what they believe a good teacher should do in the face of policies, mandates or

institutional norms. The source of the problem is dissonance between educators' moral centers and the conditions in which they work" (p. 43). The factors that contribute to teacher attrition are well known and well documented (large class sizes, low pay, under-funded schools), and in part it is the reason why high teacher turnover is a problem (teacher shortages force schools to hire inexperienced or unqualified teachers) and the kinds of educational institutions most critically affected by teacher attrition are urban and rural schools in poorer areas serving racialized groups (Carver-Thomas & Darling-Hamond, 2017). As a result, it is incumbent upon us as transformative educational leaders to address these concerns and endeavor to retain quality teachers within the educational system.

Chapter Summary

The study culminated with ascertaining a better understanding of how the output variables of the teacher instructional evaluation ratings and the assessment rubric tool that are utilized in K-12 public schools in the State of Florida are connected to input variables from the Florida Accountability Programs (FSRP and the FDAP) of FSA (R%H and M%H) and DSGs that are commensurate with an approximation to VAM/LGM scores and its impact upon the merit-based pay model that is currently implemented in the state. Merit-based pay rewards individual teachers, groups of teachers, or schools on any number of factors, including student performance, classroom observations and teacher portfolios. Merit-based pay is a reward system that hinges on student outcomes attributed to a particular teacher or group of teachers rather than on "inputs" such as skills or knowledge (Podgursky & Springer, 2007). The variables in this study were analyzed in

order to determine if this current system's methodology is constructive to generating Highly Effective educators in K-12 public schools.

CHAPTER 2. REVIEW OF LITERATURE

Bloomquist (2020) asserts that within a paideia (the rearing and education of ideal members of the State), the emphasis on logos is crucial. “It is from logos that a civilization arises and maintains itself; logos ensures that human beings can raise their spirit and aspire to reach high ethical ends” (Ballacci, 2018, p. 16). Kennedy (1994) summarizes logos-centered paideia by focusing upon getting students “to practice themes about patriotism and virtue, justice and temperance, and courage and wisdom. It is essential to have students study examples from history and choose from among these examples to illustrate their arguments” (p. 39). In this framework, students are encouraged to strive for their ambitions to be great learners, mold their character, and apply the lessons they learn. “They will try to live up to these standards, knowing that their effectiveness with an audience will result in large part from the audience’s trust in their character, the more ambitious they are, the more virtuous they will become” (Kennedy, 1994, p. 48). According to Parlak (2011):

> Accountability is a tool that ensures educational organizations have appropriate conduct in line with the law and its regulations during the administration of organizational goals. This tool also indicates a social relationship in which the administrators feel required to answer to higher authorities regarding the accuracy of their actions. In this relationship, a higher authority questions the validity of their actions and the sufficiency of information while the party that provides accountability has to answer these questions. (p. 7)

Therefore, at this juncture, an overview of the contextual history of accountability in K-12 public schools will be explored.

Callahan's (1964) premise was embedded in the notion that the educational system was being transformed into a business industrial model that many could argue is still in place in modern society. According to Callahan (1964):

> At the turn of the 20th century, America had reason to be proud of the educational progress it had made. The dream of equality of educational opportunity had been partly realized. Any white American with ability and a willingness to work could get a good education and even professional training. The schools were very far from perfect, of course; teachers were inadequately prepared, classrooms were overcrowded, school buildings and equipment were inadequate, and the education of Negroes had been neglected. But the basic institutional framework for a noble conception of education had been created. Free public schools, from the kindergarten through the university, had been established. (p. 6)

Brown vs. Board of Education

Experiences in schools are impacted by high stakes testing almost daily in the form of test preparation curriculum, benchmark assessments, or the exams themselves. "Yet, those involved have little voice in robust scholarly conversations about the impact of high stakes testing on pedagogy" (DeJaynes et al., 2020, p. 287). In *Brown v. Board of Education,* Chief Justice Earl Warren proclaimed the importance of education (Chemerinsky, 2004).

He wrote:

> Today, education is perhaps the most important function of state and local governments. Compulsory school attendance laws and the great expenditures for education both demonstrate our recognition of the importance of education to our democratic society. It is required in the performance of our most basic public responsibilities, even service in the armed forces. It is the very foundation of good citizenship. Today it is a principal instrument in awakening the child to cultural values, in preparing him for later professional training, and in helping him to adjust normally to his environment. In these days, it is doubtful that any child may reasonably be expected to succeed in life if he is denied the opportunity of an education. (Warren, 1954, p. 483)

Brown offered the promise that the federal courts would recognize a fundamental right to education and "use the Constitution to ensure equal educational opportunity for all children in the United States" (Chemerinsky, 2004. p. 115). When President Lyndon Johnson signed the Elementary and Secondary Education Act into law on April 11, 1965, even though the federal aid through the ESEA and its many reauthorizations have "contributed less than 10 percent of school funding on average, federal regulations, priorities, and evaluations attached to this aid have had a strong influence on state and local policy" (Nelson, 2016, p. 359). According to Nelson (2016) ESEA targeted poverty, but "American society is even more unequal today, both at the local and the interstate level. It bolstered state power over schools, but it also imposed significant financial and political costs that have resulted in various forms of backlash" (p. 360).

Elementary & Secondary Education Act of 1965

The Elementary and Secondary Education Act (ESEA) is the primary source of federal aid to K-12 public education. Title I was enacted in 1965 as part of the "War on Poverty". The program was intended to address a national problem that was reflected in men being rejected by the military draft, employment and manpower retraining problems, low levels of education for many adults, high unemployment rates for 18- to 24-year-olds, and "concerns expressed by institutions of higher education and vocational and technical educators regarding the quality of elementary and secondary education" (Skinner & Rosenstiel, 2018, p. 7). The ESEA was last re-authorized by the Every Student Succeeds Act in 2015 (Skinner & Kuenzi, 2015). The Title I program has always been the largest grant program authorized under the ESEA and was funded at $15.5 billion for FY2017. According to Skinner & Rosenstiel (2018), since its enactment in 1965, Title I grants provide supplementary educational and related services to low-achieving and other students attending K-12 public elementary and secondary schools with relatively high concentrations of poverty. In recent years, Title I has also become a vehicle to which a number of requirements affecting broad aspects of K-12 public education for all students have been attached as conditions for receiving Title I grants. Under current law, two formula child weighting scales are used in the determination of grants under the Targeted Grant and Education Finance Incentive Grants (EFIG) formulas. One is based on formula child rates (determined by dividing a Local Education Agency (LEA)'s number of formula children by the number of children ages 5-17 residing in the LEA, the other is based on formula child counts. Higher weights are applied to the LEAs with the highest formula child rates than are applied to the LEAs

with the highest formula child counts. “As intended, these weighting schemes would appear to favor LEAs with higher formula child rates (often rural LEAs) over LEAs with higher numbers of formula children (typically urban LEAs)” (Skinner & Rosenstiel, 2018, p. 9).

Powell Amendment

Sanders (2016) reported that Adam Clayton Powell, Jr. (D-NY), an African-American Baptist pastor first elected to the U.S. House of Representatives in 1944, “linked the federal government’s involvement in local schools to improved educational opportunities for African Americans” (p. 361). Powell created a “nondiscrimination amendment or rider that he attached to various bills over the next twenty years that prohibited the use of federal funds in support of services or programs that practiced racial discrimination” (Sanders, 2016, p. 363). The Powell Amendment became Title VI of the 1964 Civil Rights Act, which stated, “No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving federal financial assistance” (Civil Rights Act of 1964). Moreover, Sanders (2016) states, “with respect to public education in the South, this amendment only had merit if there was bait to dangle in front of segregationists. The ESEA, with its $1 billion dollars for public education, was that bait” (p. 366). The federal government was using its monetary resources to ensure that southern states which already received nearly $300 million in federal funds for vocational education, defense education programs, and the “impacted areas” programs for school districts with large numbers of government employees would receive additional funds. The ESEA doubled that amount (Spivak, 2016). In terms of

additional involvement in educational policy, "Congress instituted regulations pursuant to Title I of the ESEA that required parental involvement in schools that received these monetary funds" (Spivak, 2016, p. 4).

However, a study by Orfield and Gordon (2001) carefully documents that during the 1990s America's public schools became substantially more segregated. According to Caballero et al. (2007), educational policies and practices in schools challenge or support stereotypical perceptions of mixed heritage pupils that act as educational barriers to achievement. "The specific barriers of achievement experienced by mixed heritage pupils operate in a context where their mixed identities are absent or barely present in the curriculum as well as school, policies, and practices" (Caballero et al., 2007, p. 345). This invisibility consequently makes it difficult for the needs of pupils from mixed racial and ethnic backgrounds to be met. According to Weis and Fine (2005), "we need to understand not only the oppressive nature of policies and practices that silence, and listen closely to the words, critiques, and dreams of those who have dwelt historically on the margins" (p. 261).

A Nation At Risk Report

In addition to the various amendments and riders connected to ESEA, in April 1983 under President Ronald Regan, the U.S. DOE's National Commission on Excellence in Education released the *A Nation at Risk* (ANAR) report. It identified a crisis in the American educational system and underscored the perceived failures. This report sparked a new "standards-based reform movement" to improve student achievement (Carnoy & Loeb, 2002). This landmark report lamented the condition of American public education (Bracey, 2008). Furthermore, Bracey (2008) stated that the

commissioners assembling the ANAR report had nine national assessment trend lines (three ages by three subjects), only one of which could support crisis rhetoric. Nevertheless, “it was only that trend which appeared in the report. They used similar selectivity for other statistics and contentions. Although the report’s conclusions rested on quicksand, it launched an unprecedented level of school bashing that continues unabated today” (Bracey, 2008, p. 82).

Improving America’s Schools Act of 1994

Under President Bill Clinton, Congress then passed the Improving America's Schools Act of 1994 (IASA) which reauthorized Title I and made significant changes to its parent involvement section (Apling, 1998). “Title I is the mainstay of federal education policy and its influence is broad-sweeping” (Wells, 1994, p. 1308). Johnson (1997) underscores that in light of educational research and practice, IASA mistakenly emphasized giving parents advisory roles in school decision-making at the expense of helping schools involve parents directly in their own children's learning through parent-child learning activities (IASA, 1994). As a result, “the law is unlikely to encourage increased parent involvement or raise academic achievement” (Johnson, 1997, p. 1759). Comer (1976) suggests that “discontinuity between home and school is not caused by economic and cultural differences between parents and teachers” (p. 540). Rather, it is caused by “organizational arrangements that impede cooperation and perpetuate teachers' and parents' stereotypes, misperceptions, and lack of understanding of mutual needs” (Leitch & Tangri, 1988, p. 72). Cochran and Dean (1991) state that “schools are key institutions in local communities and thus are in a special position to enhance or retard the empowerment process” (p. 261).

Another scholar declared:

> For a long time, we have understood that the magic of suburban schools is not merely the relative affluence and abundant resources of the citizens, but also the balance of power between families and schools, the sense of responsibility and accountability teachers feel for the educational success of children, and the parent's sense of entitlement in demanding results from schools. (Lightfoot, 1981, p. 101)

The ESEA has been subject to several comprehensive amendments, of which "the IASA is one of the most dramatic" (Johnson, 1997, p. 1777). When ESEA's budget allocation was due to expire in 1993, the Clinton Administration proposed significant new directions for the program, the bulk of which were included in the final version of the IASA bill, which was signed on October 20, 1994 (Hawkins-Stafford Amendments, 1994). Both Congress and the Department of Education sought to use ESEA's Title I component for parent involvement provisions to raise academic achievement among the target student population. However, the "IASA neglected parent impact techniques as tools for reaching this goal" (Johnson, 1997, p. 1785).

No Child Left Behind Act of 2001

Under the George W. Bush Administration, reauthorization of ESEA was now known as the No Child Left Behind (NCLB) Act of 2001 which intended to ensure that all children receive a high quality education and "close the achievement gap between high and low performing children, especially the achievement gaps between minority and non-minority students, and between disadvantaged children and their more advantaged peers" (Elementary Secondary Education Act, 2015). Prior to NCLB, many states

channeled their enthusiasm for standards-based education towards creating local assessment policies (Berliner, 2006). With the passage of the NCLB (2002), the early 2000s throughout the U.S. marked a new era in public educational accountability policies, with federal policies increasingly promoting accountability-based systems that held students, teachers, and schools responsible for improved student achievement results. Some research indicated that teachers affected student performance and that teacher performance differed within schools (Rivkin et al., 2005; Rockoff, 2004). Despite this, most teacher evaluation systems are based primarily on principal and administrator observations which indicated that almost all teachers received satisfactory results (Weisberg et al., 2009). In response to wide variation in teacher quality and persistent student achievement gaps among student subpopulations, policymakers have considered a number of alternatives, including increasing the monetary value placed upon teacher evaluations (i.e. Marzano iObservations ®, FSA – English in the domain of Reading [R%H], Mathematics [M%H]), and District School Grades (DSGs) which are commensurate with an approximation to VAM/ LGM scores and accountability (Ingle, 2015), and the application of free-market and economic theories to schooling and teacher compensation reform (Podgursky & Springer, 2007, 2009).

According to Jahng (2011), "the test-based accountability system established single performance goals for minority and non-minority children that they are expected to meet; it also requires that schools make adequate yearly progress (AYP) on state reading and mathematics tests" (p. 100). By expanding the federal control over state and local systems, the federal government requires public schools to report the test results students obtain in reading and math (Sunderman et al., 2005). Moreover, if a school does not meet

the performance standard given by NCLB, the school is labeled a "failing school" and "punished through withdrawal of federal funds, pressure for privatization, and public school choice" (Fusarelli, 2004, p. 72). The NCLB policy focused upon accountability for states, school districts, and schools; greater choice for parents and students; more flexibility for states and local educational agencies in the use of federal education funds; and a strong emphasis on reading for young children (U.S. DOE, 2011). NCLB required states "to adopt content and achievement standards, to measure student progress toward those standards, and to implement a series of interventions and sanctions in schools and districts that fail to meet their targets" (Hamilton et al., 2007, p. 1). With the passage of NCLB, "a much larger accountability role for the federal government was specified, including requiring public schools and states to report student achievement and be subjected to federal standards" (Brookhart, 2009, p. 2). However, Brookhart (2009) asserts that many researchers now agree that NCLB did not meet its intended effects of attaining 100% student mastery of higher standards by 2014. More specifically, research suggests that since the passage of NCLB, many students, especially those in the country's lowest performing public schools, have been increasingly susceptible to unprofessional test-based practices including: teaching to the tests, teaching test preparation, test practice, and test rehearsals instead of curricular content; teaching while hyper-emphasizing the rote memorization of facts and basic skills likely to be on tests; narrowing the curriculum to match the content and concept areas tested; and, related, teaching the tested subject areas that "count" (i.e., mathematics and reading/language arts) while marginalizing or even eliminating other curricular areas and activities that do not "count" on high-stakes tests (i.e. art, music, and physical education) (Amrein &

Berliner, 2002; Haney, 2000; Nichols & Berliner, 2008). Despite mandating that all states assess students, NCLB did not provide a nationalized system of standards or a systematic method of assessment. This allowed for state-level variation in standards and testing policy (Wenning et al., 2003).

Considered a state at the forefront of the standards and accountability movement (Berliner, 2006), Virginia created the Standards of Learning (SOL) in 1995. Approved by the Virginia Board of Education, the first iteration of the SOL encompassed four main content areas of math, science, English, and social studies and assessed students throughout primary and secondary school years (Weisberg et. al., 2009). According to Berliner (2006), each SOL examination included a corresponding curriculum framework that public school teachers used to ensure proper lesson planning and content coverage. Berliner (2006) also asserts that "under the SOLs, failing public schools were subject to academic reviews and were required to submit school improvement plans, including new teaching programs grounded in research in student achievement" (p. 23). Failing these changes, schools could be closed, combined with a more successful school, or be reconstituted which included complete school restructuring. Other states, such as Florida, utilized a similar model for assessment. Carnoy and Loeb (2002) found that this high-stakes testing system allowed critical educational decisions to be made for students and teachers based on results. "The SOLs were found to place considerably high external pressure on schools, students, and teachers based on an external assessment of test frequency, accountability, and repercussions of poor performance" (Carnoy & Loeb, 2002, p. 316). In 2010, the standards movement continued with the adoption of the CCSS. Created to address the lack of standardization in state learning standards and the

definition of proficiency, “the CCSS received both praise and criticism from experts and policymakers alike from across the United States” (Ruff, 2019, p. 4).

Every Student Succeeds Act

In 2015, NCLB was replaced under President Barack Obama by the Every Student Succeeds Act (ESSA). Ruff (2019) asserts that as a bipartisan congressional effort, ESSA was viewed as a return to normalcy, correcting the previous overreach of the federal government into state education policy. In announcing ESSA, the U.S. DOE recognized that the prescriptive requirements of NCLB had become increasingly untenable (ESSA, 2018). While still maintaining the annual standardized testing requirement as established by NCLB, ESSA shifted other previously-held federal accountability provisions to the states. In particular, ESSA restored a considerable level of control over standards and testing to states and districts, which would now be required to submit goals and standards to the U.S. DOE for approval. In this way, “ESSA wrestled power away from the U.S. Department of Education which had under Secretary Arne Duncan assumed oversight of many of the federal lawmaking procedures” (Ruff, 2019, p. 5). Additionally, President Obama’s Administration simultaneously instituted the RTTP program, which provided a series of financial incentives for states to tie test scores to teacher evaluations, adopt the CCSS, and institute sweeping and prescriptive school turnaround strategies (Committee on Health, Education, Labor, and Pensions, 2010). “It sparked a roiling debate over states versus federal rights” (Burnette, 2016, p. 2). Burnette (2016) asserted that parents and voters have deep anxieties about a shifting economy, the resegregation of schools, and the widening achievement gap between states’ growing minority population and its wealthier students, “all policies states will have to confront

and deal with its ESSA and RTTP plans in the future" (p. 7). In Florida, however, Governor Ron DeSantis signed an executive order to end the use of the CCSS. The State's Education Department had to make recommendations on how to eliminate those standards "and ensure we return to the basics of reading, writing, and arithmetic" (Sawchuk, 2019, p. 7). According to Sawchuk (2019), the standards ultimately became a political football in Florida, especially after the U.S. DOE encouraged their adoption. "Florida already made one batch of revisions to the standards in 2014, and rebranded them to the Florida Standards Assessment (FSA). However, they remain substantially similar to the common core" (Sawchuk, 2019, p. 3).

On February 7, 2017, under President Donald Trump, Betsy DeVos, a billionaire businesswoman, Republican megadonor and long-time school choice advocate, became steward of the nation's nearly 100,000 public schools (Levy, 2017). "DeVos is the first Education Secretary in U.S. history who has not been a government official, school administrator, teacher, public school parent or student" (Kaplan & Owings, 2018, p. 58). During her confirmation hearings, she expressed confusion over federal law and education policy, refused to commit to maintaining funding for public schools, and would not agree with the need for equity and accountability for all federally funded schools (Levy, 2017). The U.S. DOE has a $70 billion dollar budget and federal oversight of the nation's 98,000 public schools that educate around 90% of America's children (National Center for Education Statistics, 2016). DeVos champions a market-based privatization of public education using charter schools and vouchers. Her pro–school choice advocacy group, the American Federation for Children (2017), works to create programs and pass laws that require using public funds to pay for private school tuition via vouchers and

similar programs. Yet, her success in spreading charter schools in Michigan, her home state, has not been matched by student achievement; most Michigan charter schools have recorded student achievement test scores in reading and math below the state average (Emma et. al., 2016; Joy & Arellano, 2016; Lenhoff et al., 2013). Increasing education delivery options and privatization—not successful student learning—appeared to be Secretary DeVos' primary focus. In her Senate hearings, DeVos compared school choice to taking Uber or Lyft over a taxi, portraying education as a business responsive to market forces of consumer choice and innovation through competition (Brookings Institution, 2017).

In 2020, President Joe Biden nominated Dr. Miguel Cardona, Connecticut's first Latino Education Commissioner, to serve as U.S. Secretary of Education. The Senate Committee on Health, Education, Labor, and Pensions confirmed his appointment in March 2021. Cardona's experience lies in the K-12 sphere. Nevertheless, Biden described Cardona as an "innovative leader" who would "eliminate long-standing inequities, close racial and socioeconomic opportunity gaps and expand access to community colleges, training and public four-year colleges and universities" (Weissman, 2021, p. 5). It is yet to be comprehensively determined how President Biden's selection for Education Secretary, Dr. Miguel Cardona, a career educator, will fare in the age of Covid-19, the worst pandemic health crisis in a century.

National Defense Education Act of 1958

At no time in the history of U.S. public schools have those responsible for schools been unaccountable. According to Sirotnik (2004), with the early 19th-century stirrings of tax-supported public education, "the state-chartered district school board with elected

trustees was legally obligated to the local community to ensure that children who attended public school were adequately housed, taught, and had materials to achieve the community's goals for its public schools" (p. 19). President Dwight D. Eisenhower signed the National Defense Education Act in 1958. A cascade of reforms raised graduation requirements in math and science, added programs for the gifted, and introduced advanced placement (AP) high school courses to speed entry into colleges (Strain, 2005).

Educational Accountability Act of 1971

In Florida, the Educational Accountability Act of 1971 (EAA) was enacted to provide for the implementation and further development of public school education assessment procedures as required by the laws of the State (EAA, 2021). The plan for educational assessments in Florida which was developed by the Commissioner of Education provided for the establishment of educational accountability in the public education system of Florida to assure that education programs operated in the public schools led to the attainment of established objectives for education. "The focus of federal and state policymakers shifted from concerns about racial segregation as an obstacle to achieving educational equality to an emphasis on accountability as a mechanism to enhance academic achievement" (Borman et al., 2004, p. 608). Moreover, it provided information for accurate analysis of the costs associated with public education programs and information for an analysis of the differential effectiveness of instructional programs (EAA, 2021).

Educational Reform Act of 1983

The Educational Reform Act of 1983 required an annual report by K-12 public school districts that focused upon the continuation of education programs, including pupil services, cost effectiveness, attendance patterns, dropout rates, graduation capabilities, and involvement with occupational training. “Merit pay was proposed as the panacea for the ailing American educational system that would enable it to catch up with and surpass the academic achievements of students from other countries” (Arthur & Milton, 1991, p. 266). Reports had to be filed annually and the Superintendent of Public Instruction had to summarize and evaluate annual reports from the K-12 public school districts by the beginning of December of each academic school year (Kasper, 2005).

Blueprint 2000

In 1991, the Florida State Legislature approved Senate Bill 2054, the Education and School Improvement Act (ESIA) otherwise known as "Blueprint 2000” which was “a plan for K-12 public school improvement and accountability” (Terzian & Boyd, 2004, p. 135). Blueprint 2000 was created to improve the performance of K-12 public school students and educational programs by returning the responsibility to those closest to the students. The cadre consisted of the schools, teachers, and parents. The Office of Program Policy Analysis and Government Accountability (OPPAGA) reported that by the year 2000, the Legislature intended for Florida to establish a system of school improvement and educational accountability based on the performance of students and educational programs (OPPAGA, 1996). According to Terzian and Boyd (2004), while the overall purpose of Blueprint 2000 was to improve the performance of the K-12 public school students, the process of shifting responsibility for improvement to local

communities affected other areas such as the involvement of stakeholders in the school improvement process, the allocation of financial resources, and the decision-making at the local level.

In Florida, K-12 public school teachers and principals were surveyed for three school years (1993-94;1994-95;1995-96) by the Office of Program Policy and Government Accountability (OPPAGA). The prevailing belief was that by implementing various school improvement initiatives their schools were making improvements relative to the seven state education goal areas. Although teachers and principals believed their schools were making improvements related to all seven state education goals, they most often said they were improving student performance (OPPAGA, 1996).

A+ Plan for Education

Borman and Dorn (2007) assert that standards, testing, and accountability represent the infrastructural elements of contemporary market-inspired reforms implemented at the federal and state levels. In Florida, the standards-based accountability reform policy is called the A+ Plan. The A+ Plan was designed to address both accountability and improvements to student learning. According to the Center for Education Reform (2019), "in 1999, more than 60% of minority and low-income fourth graders in Florida could not read at a basic level. Additionally, barely half of Florida's high-school seniors were graduating" (p. 2). The A+ Plan provided professional development for K-12 public school teachers, high academic standards, school safety improvements, and student truancy reductions. As of 2007, close to $900 million in additional funding was allocated for K-12 public schools (Borman & Dorn, 2007).

Florida School Recognition Program

At this juncture, the State of Florida maintains school accountability reports. According to the Florida Grading System Act which can be found in Florida Statute § 1008.34 (2021), this new incarnation of accountability is known as the FSRP which determines that there is a need for a performance incentive program for outstanding K-12 public school faculty and staff in highly productive schools. The Legislature further found that performance-based incentives are commonplace in the private sector and should be infused into the public sector as a reward for productivity (Florida State Legislature, 2021). Bergin et al. (2017) assert that some teachers could be advantaged or disadvantaged by differences in rater accuracy related to idiosyncrasies in rater judgments, including biases and erroneous applications of the scoring rubric. Moreover, "some teaching episodes and some teaching practices (i.e. the use of student and personnel assessments) were more difficult to rate accurately" (Bergin et. al., 2017, p. 26). The FSRP is created to provide financial awards to K-12 public schools that either:

a) Sustains high performance by receiving a school grade of "A," making Excellent Progress. (FLDOE, 2020a)
b) Demonstrates exemplary improvement due to innovation and effort by improving at least one letter grade or by improving more than one letter grade and sustaining the improvement the following school year. (FLDOE, 2020a)

All K-12 public schools, including charter schools, that receive a school grade pursuant to Florida Statute §1008.34 are eligible to participate in the program (FLDOE, 2021a). With this type of incentive accountability measures in place, it begs the question of the

ramifications of this business model upon merit pay salary schedules for K-12 public school educators.

Florida Differentiated Accountability Program

The Florida Grading System Act (2021) serves as the foundation of the Florida Statute § 1008.34 section which establishes the differentiated accountability of state support for school improvement, also known as DA, in which graded, non-charter schools and their districts are identified for interventions, support and monitoring based on their school grade history (FLDOE, 2016). The FDAP combines the federal and state accountability programs. It was created to help schools increase student achievement and implement successful academic improvements (FLDOE, 2016). The FDAP and the FSRP are interconnected in that accountability and financial remuneration are linked within the context of K-12 public education.

Merit Pay

In 1983, the Committee on Education and Labor in the House of Representatives in the first session of the 98th Congress prepared the Merit Pay Task Force Report. It professed the notion that:

> The American school system is an expression of the value we hold for education for all who wish to avail themselves of it and it provides a social process of opportunity for all children. But public and governmental concern often has been like a roller coaster ride: sudden ascents and even more rapid declines. (p. 4)

It expressed a fundamental belief that there were problems in the public school system and if changes were not made then the future of the county would be vulnerable and insecure. The Merit Pay Task Force Report (1983) proffered that **"**the one essential

ingredient for a superior educational opportunity is a talented, dedicated teacher. Yet, as a nation, we pay teachers less than most professionals" (p. 5). Merit pay is arguably the most popular form of incentive pay used by organizations, although evidence concerning its success in practice remains mixed (Gerhart et al., 2003; Mitra et al., 1997).

Historical Theory of Teacher Merit Pay

The theory of teacher merit pay has been well canvassed and deconstructed (Marsden et al., 2007; Podgursky & Springer, 2007). Merit pay is expected to operate by providing stronger incentives for teachers to work harder and by encouraging more effective individuals to select into the teaching profession. Lazear (2004) argues that in the case of teacher merit pay, selection effects may be more important than incentive effects. All performance pay schemes are subject to criticism on the basis of unfairness. Bewley (2004) asserts that in an extreme case, a student's performance may be affected by factors outside the school's control. Such factors might include family inputs (i.e. parental education or the amount of time parents spent reading with the child) or idiosyncratic shocks (i.e. whether a student fight occurs outside the classroom on the day of a test). In this scenario, teachers cannot affect students' performance, so the effect of merit pay is merely to introduce random variation into pay rates. To the extent that teachers place a high value on the fairness of their remuneration system, "merit pay may end up reducing morale and effort" (Hoxby, 2002, p. 21). Test-based merit pay schemes are also criticized on the basis that they may cause teachers to focus on a narrow subset of activities, or to "teach to the test" (Leigh, 2013). Campbell (1979) definitively maintained that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to

distort and corrupt the social process it is intended to monitor" (p. 71). Jacob and Lefgren (2008) contend that merit pay schemes that use principal and assistant principal ratings are criticized on the basis that they are vulnerable to favoritism. In numerous school settings, it is argued that many school principals and assistant principals "do not have the skills to administer the pay of the teachers in their schools and it is better that they be left to focus on educational leadership without having to also worry about salary determination" (Jacob & Lefgren, 2008, p. 105).

Another concern of merit pay schemes is that they are based on individual incentives which may reduce the incentive for teachers to collaborate with one another. Leigh (2013) maintains that compared to some other occupations, teaching involves relatively little team work. However, Perez (2011) contends that there are still instances in which teachers assist one another, through team-teaching, sharing classroom notes, or providing mentoring and advice in the staff room and planning area. Moreover, "some experimental evidence suggests that teachers may be more averse to inequality than other professionals" (Perez, 2011, p. 7).

Furthermore, Leigh (2013) posits that "merit pay schemes that are based on group incentives (i.e. schemes that provide a reward if the entire school does well) are not productive on the basis that they are vulnerable to a free-rider problem" (p. 4). For example, suppose that teachers in a school are told that they will all receive merit pay if their students' test scores exceed a particular threshold. "Each teacher will therefore know that his or her own effort will only make up a small portion of the teaching staff's contribution to the probability that the school meets that threshold" (Neal & Springer, 2009, p. 158).

The work of Chubb and Moe (1990) argue that problems of academic performance result from the regulation of schools by public bureaucrats who respond to the interests of organized groups and not to the interests of students. According to Smith et al. (1993), performance standards, curriculum, and student assessment are not purposefully integrated and federal and state policies often work at cross purposes. In Florida, test-based incentives include attempts to increase school accountability through the regular testing of K-12 public school students, making the results, aggregated to the school level – public, and rewarding schools with high or increasing aggregate test scores, and imposing sanctions on poor performing schools (Rouse et al., 2013).

Merit Pay in the Business Realm

Kanter (1987) states that "status not contribution has traditionally been the basis for the numbers on employees' paychecks. Pay has reflected where jobs rank in the corporation hierarchy – not what comes out of them" (p. 60). Despite centuries of experience with employee compensation plans, status still wins out over contributions. "Contrast the pay of corporate CEOs in the U.S., some of whom earn more than 600 times as much as their non-management employees, to the much smaller differential between the pay of employees and CEOs in Europe and Japan" (Ramirez, 2001, p. 16). In turn, the seemingly logical link between employee production and compensation is often highly subjective. Although merit pay and bonuses for managers are common forms of compensation, there have been no rigorous tests of their effectiveness (Heneman & Schwab, 1978). Deci and Ryan (2010) conducted a series of studies on the effects of externally mediated rewards, such as pay, on laboratory subjects' intrinsic motivation to engage in tasks. Deci and Ryan (2010) drew on this research to argue that "contingent

payment plans should be avoided because they reduce intrinsic motivation, lead individuals to develop strategies that will enable them to get rewards with the least effort and can easily break down" (p. 2). Pearce et al. (1985) assert that "the theory and research tying compensation to organizational performance shows there is a lack of conclusive empirical support for this assumption" (p. 261).

It is important to question whether effective assessment practices persist in a performance pay regime. Indeed, a primary reason the single salary schedule replaced the grade-based compensation system was that subjective measures used to reward teachers were highly susceptible to gender and racial discrimination as well as nepotism (Marsden et al., 2007). Hence, the variables that can impact merit-based pay programs reflect the complexity involved in monitoring teacher performance. According to Murnane and Cohen (1986), teacher performance is more difficult to monitor than performance in many other professions because output is not readily measured in a reliable, valid, and fair manner. Thus, it is argued that the education sector cannot readily measure the value of the services provided by an individual teacher or group of teachers, since achievement can oscillate from a constellation of factors beyond the instructor's control.

The Impact of Merit-Based Pay in Public Education

Merit-Based compensation schemes are rooted in behavioral psychology. Skinner et al. (1988) contend that its basic premise is that the behavior of all organisms can be explained entirely by the desire to maximize rewards and minimize punishment. Heneman (2002) also maintains that the intuitive appeal of this argument is so great that it now seems a matter of common sense to motivate people to act in a specific way and reward them whenever they do so. "To encourage people to cease specific behaviors,

threaten to punish them. Rational beings will work to maximize rewards and avoid punishment" (Heneman, 2002, p. 12). There is no way to determine clearly whether a specific behavior is driven by extrinsic or intrinsic motivation (or a combination of both). In practice, "it is not necessary to be certain of the nature of people's motivations to understand how a merit-based compensation system rooted solely in extrinsic rewards would be detrimental to the overall health of an organization and its members" (Benton & Radziwill, 2016, p. 15).

Ryu and Jinnai (2020) assert that "public schools introduced a school-based merit pay program in which teachers from high-performing schools received monetary incentives regardless of individual teachers' quality" (p. 206). Public schools nationwide implemented carefully designed school accountability programs. The accountability program introduced (a) statewide standardized tests, (b) an accountability system for each school's average test scores, and (c) monetary incentives for teachers receiving highly effective personnel rating results at high-achieving schools. All students in Florida K-12 public schools are required to take standardized FSA tests for students in Grades 3 through 8 in Math and Reading as well as for those in Grades 9 through 12 at the end of every school year (Rouse et al., 2013). Additionally, Ballou and Podgursky (1993) state that "teacher opposition is a fundamental reason for the failure of merit pay plans and that teachers must be involved in all stages of program design as well as implementation if pay-for-performance is to succeed" (p. 50).

Department of Education Analysis of the Florida School Recognition Program

The FSRP was created by the Florida Legislature in 1997 to provide a financial incentive to schools and teachers that achieve specific standards-based goals.

Achievement is measured primarily through the state's current high-stakes test, the FSA. By framing public schooling outcomes in terms of the "productivity" to be found in the "private sector" (FLDOE, 2004), three assertions about the nature of K-12 public schooling in the State of Florida became clear:

1. Public schooling is a competitive process, with schools serving as the competitors.
2. It is the responsibility of schools to produce measurable outcomes.
3. Higher rates of production deserve increased reward. (FLDOE, 2004)

The language of the statute best represents the intent of the FSRP: The Legislature finds that there is a need for a performance incentive program for outstanding K-12 public school faculty and staff in highly productive schools. The Legislature further finds that performance-based incentives are commonplace in the private sector and should be infused into the public sector as a reward for productivity (FLDOE, 2004). The FSRP was specifically established to:

> Provide financial awards to K-12 public schools that: (a) Sustain high performance by receiving a school grade of "A," making excellent progress; or (b) Demonstrate exemplary improvement due to innovation and effort by improving a letter grade (c) All public schools that receive a school grade are eligible to participate in the program. (FLDOE, 2004)

The FSRP was part of Florida's entry into the "high-impact" (Malen, 2003), standards-based reform movement. Anointed as the "Sunshine State Standards," the original format was chosen to provide flexibility to K-12 public school districts in designing curriculum based on local needs (FLDOE, 2004). However, as Florida moved

toward greater accountability for student achievement at each grade level, the Sunshine State Standards have been further defined (FLDOE, 2004). “The Sunshine State Standards provided an increasingly defined reference point for the productivity to which the “incentives” set forth by the FSRP are attached” (Gayles, 2007, p. 440). Despite this language, “the Florida legislature has appropriated comparatively few dollars to the program designated to close the achievement gap” (Gayles, 2007, p. 443). As accountability to high standards is implemented in Florida, the overwhelming weight of fiscal emphasis is placed on those K-12 public schools that are performing “to standard” and “above standard,” whereas schools that are not able to perform in this manner, for whatever reason, are largely ignored fiscally (FLDOE, 2004). Furthermore, while FSRP funds are disbursed at each individual school’s discretion, Differentiated Accountability (DA) and Assistance Plus Program dollars are disbursed to low-performing public schools through facilitators, coaches, and community oversight boards” (FLDOE, 2004).

Competition, as it plays out in K-12 public schools in Florida and to an increasing degree the rest of the nation, primarily means high-stakes testing is attached to state standards (American Federation of Teachers, 1999). Despite the increased reliance on testing in K-12 public schooling, a growing body of literature asserts that high-stakes testing expands the “achievement gap” rather than reduces it (Anyon, 1995; Darling-Hammond, 2004; Diamond & Spillane, 2004; Gillborn & Youdell, 2009; Lipman, 2002; Lipman & Gustein, 2001; McNeil, 2000; Sandholtz et al., 2004; Scheurich et al., 2000). The allure of testing within school-based reform reveals the empty dream of open competition and meritocracy in schooling (Attanucci, 2004; Bowles & Gintis, 2011; Goldthorpe, 2003; Scully, 2002). The implementation of high-stakes testing in K-12

public schools does not occur in isolation. When examining the concept of American meritocracy which is the belief that individuals earn what they receive on the basis of the merit of their individual performance, Gayles (2007) contends that "this powerfully idealistic construct is denigrated in the presence of outcomes that are marred by disproportion and disparateness" (p. 449). According to Gayles (2007), the concept of meritocracy is inextricably bound to notions of fair play. When resources are limited, competition is required if the allocation of these resources is to be considered meritocratic. "By invoking the private sector and productivity, the Florida legislature pitted schools against one another for a conceptually limited resource: the FSRP dollars" (Gayles, 2007, p. 451). In Florida, the role that FSRP and FDAP play in educational policy serve to perpetuate the differentiated stratification (Gayle, 2007) that exists at multiple levels of a school ranging from the teachers to the students that they serve.

Merit-Based K-12 Public School Pay-for-Performance Schemes in Florida

In Florida, proponents of merit pay point out that teaching has considerably more pay compression than most occupations. "While intra-occupational salary dispersion has risen in most developed nations, it has remained relatively unchanged within teaching" (Leigh & Ryan, 2008, p. 142). A similar pattern can be seen when following the same individuals over time. For example, Chingos and West (2012) show that teachers who are more effective in the classroom tend to earn more once they leave the teaching profession—despite there being only small pay differences across those who remain in the teaching profession.

According to Brewer et al. (2015), educational reforms have become the new policy mainstay in educational discourse. Without a doubt, "fixing" teachers and

increasing public school student test scores have both been a large component of much of the reform rhetoric. Moreover, calls for implementing merit pay schemes have combined reformers' efforts to "fix" teachers while increasing test scores as public school teacher pay is linked directly to student academic achievement. Merit pay schemes situate the current push for merit pay within the education reform movement, while highlighting the overt and covert implications of injecting competition into teacher salaries. "In addition to creating an environment that lends itself to narrowed pedagogical approaches and teaching to tests (and even cheating on them), the merit pay schemes require teachers to compete with one another and may undermine positive collaboration" (Brewer et al., 2015, p. 45). Money can be a powerful motivation. The prospect of making more money often encourages individuals and corporations to work harder, smarter, and more efficiently. In this way, monetary incentives can serve as the proverbial carrot to elicit a desired reaction out of oneself or others. Similarly, the threat of losing money can also be a powerful incentive. In the role of the proverbial stick, the threat of losing money often increases self-reflection and heightens intentionality (Brewer et al., 2015). Brewer et al. (1776/2015) stated that Economist, Adam Smith, noted that "it is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own self-interest" (p. 7).

In the context of K-12 public schooling, teacher salaries have traditionally been linked to a teacher's level of post-secondary education and duration of service. However, in the present age of market-based educational reforms, merit pay schemes are becoming an often cited silver bullet to fixing the "horrid state of public education" that was first exclaimed in *A Nation at Risk* (National Commission on Excellence in Education, 1983).

This failure to investigate reforms that rest on dubious grounds has and continues to leave education susceptible to policies and policymaking that may be rooted in unethical and non-transparent motivations. Ong (2007) asserts that "a state-centric, economic strategy where optimal gains and profit are sought through strategies of self-governance encourages individuals and groups to attune themselves to the market and market-based path dependence" (p. 4). It is through this capitalistic logic, where "its social relations, its laws of motion, its contradictions—the logic of commodification, accumulation and profit maximization that penetrates every aspect of our lives" (Wood, 1997, p. 551). Brewer et al. (2015) maintain:

> The goal of commodification is the proliferation of the disposition and ideology that reinforces beliefs that a teacher's individual lesson plans, content knowledge, etc., are ultimately seen as individualistic commoditized goods that are to be hoarded within this constructed competitive environment of having "better" teachers and higher pay that may undermine collaboration. (p. 47)

What would then have to be analyzed is how teacher merit pay facilitates this monetary-based thought process and the practices that impact the public school systems.

Conceptual Framework of the Teacher Evaluation

The construct of the teaching performance evaluation is a thematic research line on a global, national, and state level. Since the role of the teacher in the school has evolved and relevant teaching-learning processes must be established, its improvement has been prioritized. Unquestionably, the teaching performance evaluation is complex and its application is diverse according to the regions and existing political systems (Rivas, 2015; Vaillant, 2016). The teaching performance is understood as the observable

pedagogical practice and it manifests itself when the teacher expresses his or her competency, and has to do with the expected learning achievements (Benitez et al., 2017). The intentionality of education and the execution of the assigned tasks depends on different factors related to quality and the initial training of teachers in order to achieve levels of excellence in education (Benitez et al., 2017).

Teacher quality has been consistently identified as the most important factor affecting student achievement (Looney, 2011; Muijs et al., 2014; Papay, 2012). This finding has spurred a nationwide movement toward the improvement of K-12 public school student achievement (Ahn, 2013). As a result, "several states have now implemented accountability measures aligned with teacher evaluation methods, which rate teacher quality and effectiveness" (Hinchey, 2010, p. 5). The Obama administration began RTTT, a federally funded grant competition, to encourage states to revamp their academic curricula in order to further encourage student growth and achievement (Harris et al., 2014; Stumbo & McWalters, 2011). The two most commonly implemented evaluation systems for evaluating teacher performance and effectiveness include the VAM/ LGM, and a standards-based teacher observation system (Looney, 2011; Muijs et al., 2014; Papay, 2012). Papay (2012) also asserted the VAM and LGM (i.e. DSGs which are commensurate with an approximation to VAM/LGM scores) evaluation method is quantitative and seen as an objective tool that is based on student achievement and growth in standardized tests compared to other students throughout the state. Alternatively, the standards-based teacher evaluation method refers to classroom observations that are subjective and can be skewed with the evaluator's biases (Papay, 2012). The scrutiny of these evaluation methods stems primarily from the accountability

measures associated with their results, which include employment-related decisions such as tenure, pay, and dismissal (Florida Personnel Evaluation Procedures and Criteria Act (2020) - Florida Statute § 1012.34; Papay, 2012). Since the stakes are high for teachers, questions emerge concerning the validity of these teacher effectiveness and accountability measuring tools, especially in regards to the more subjective teacher observation evaluation (Darling-Hammond et al., 2012).

Florida's Evaluation System Focusing Upon Approval and Reporting

Under the Florida Personnel Evaluation Procedures and Criteria Act (2020) which can be found in the Florida's Statute §1012.34 which focuses upon instructional personnel evaluation procedures and criteria, for the purpose of increasing student academic performance by improving the quality of instructional services in the K-12 public schools of the state, the district school superintendent shall establish procedures for evaluating the performance of duties and responsibilities of all instructional personnel employed by the school district. The district school superintendent shall provide instructional personnel the opportunity to review their class rosters for accuracy and to correct any mistakes. The district school superintendent shall report accurate class rosters for the purpose of calculating district and statewide student performance and annually report the evaluation results of instructional personnel to the Department of Education in addition to the information required under this subsection (Florida Personnel Evaluation Procedures and Criteria Act - Florida Statute §1012.34, 2020). Therefore, the primary power source that controls this process is placed in the purview of the Florida School Districts' School Superintendents. The transparency and oversight involved in this complex process are not necessarily consistent throughout the state.

The Florida Evaluation System Requirements for the K-12 public school teaching instructional personnel must include the following:

(a) Be designed to support effective instruction and student learning growth, and performance evaluation results must be used when developing district and school level improvement plans. (Florida Personnel Evaluation Procedures and Criteria Act [2020] - Florida's Statute §1012.34)

(b) Provide appropriate instruments, procedures, timely feedback, and criteria for continuous quality improvement of the professional skills of instructional personnel and school administrators, and performance evaluation results must be used when identifying professional development. (Florida Personnel Evaluation Procedures and Criteria Act [2020] - Florida's Statute §1012.34)

(c) Include a mechanism to examine performance data from multiple sources, including opportunities for parents to provide input into employee performance assessment evaluations when appropriate. (Florida Personnel Evaluation Procedures and Criteria Act [2020] - Florida's Statute §1012.34)

(d) Differentiate among five levels of performance as follows:

1. Highly Effective
2. Effective
3. Needs Improvement or, for instructional personnel in the first 3 years of employment who needs improvement, developing
4. 3 Years – Developing
5. Unsatisfactory. (Florida Personnel Evaluation Procedures and Criteria Act [2020] - Florida's Statute §1012.34)

K-12 Public School Teaching Instructional Personnel and school administrator performance evaluations must be based upon the performance of students assigned to their classrooms or schools, as provided in this section (Florida Personnel Evaluation Procedures and Criteria Act [2020] - Florida's Statute §1012.34). The Commissioner of Education shall approve a formula to measure individual student learning growth on the statewide, standardized assessments in English Language Arts and Mathematics that are administered annually section (Florida Personnel Evaluation Procedures and Criteria Act [2020] - Florida's Statute §1012.34).

Marzano Teacher Evaluation Model and the iObservation ® Tool

Florida Senate Bill 736, passed in 2011, rewrote how teachers are paid and retained across the state (Florida Senate, 2016). The law required districts to rate K-12 public school teachers and administrators annually (Florida Senate, 2016). In 2011, the FLDOE adopted the MTEM as the state model. However, each district was able to weight the model as they preferred in order to evaluate teachers (FLDOE, 2014). The MTEM was implemented in Florida and is utilized by K-12 public school leaders at each site. The iObservation ® rubric which is a part of the MTEM was developed by Dr. Robert Marzano and focused upon the Art and Science of Teaching. According to the Marzano Learning and Sciences Center for Teacher and Leader Evaluations (2014), the MTEM includes 41 revised elements (i.e., categories of instructional strategies) that are organized into nine broader categories which themselves are organized into three lesson segments (I. Routine Strategies, II. Content Strategies, and III. Strategies Enacted on the Spot). While in alignment with the Common Core State Standards (CCSS) followed by the FSA (2015), it was determined that seven of the 41 elements in the model should become

staples of classroom instruction (Marzano et al., 2013). At the present time, Florida still uses the MTEM as the rubric for K-12 public school teacher evaluations. States and school districts across the nation have been responding to the legislative reforms by including evaluative measurement systems for teachers (Alger, 2012; Auguste et al., 2010; Bill & Melinda Gates Foundation, 2012; Rentner & Kober, 2012).

These systems call for higher accountability for K-12 public school teachers and a focus on improved educator effectiveness and student learning. According to Basileo and Toth (2019), MTEM lacks large scale empirical investigations to assess its predictability and efficacy. "In addition to student performance, a K-12 public school teacher's IP score is created, and it includes teachers' ratings from their formal observations conducted throughout the school year" (Basileo & Toth, 2019, p. 2). The iObservation ® tool is used as a premiere metric of evaluation, therefore, the instrument needs to be examined to determine if "empathy may be instrumental not only in building bonds with those being evaluated but checking to see if it provides support for self-worth, and individual consideration, with effective communication, problem-solving, decision-making, and ultimately performance" (Kellett, 2006, p. 17).

Value-Added Model/Learning Growth Model

Value-Added Model (VAM) and Learning Growth Model (LGM) measures, sometimes referred to as growth measures, are also used to estimate how much positive or negative impact teachers have on students' achievement during a given school year. "VAMs and LGMs aim to isolate a teacher's contribution by controlling for student, classroom and school-level measures, thereby making it possible to study individual growth and compare teachers in different classrooms and schools" (Basileo & Toth,

2019, p. 6). The DSGs were used as a commensurate approximation to individual teacher VAM/LGM scores since the FLDOE does not release that data to the public. The current literature reflects that observation scores collected from the evaluation framework of the MTEM, which is widely used in Florida, has received little attention in the scientific community in regards to its ability to reliably predict teacher value-added measures (Basileo et al., 2015). In Florida, K-12 public school state FSA assessment data (R%H, M%H, and DSGs) are the basis for the value-added models. The models measure the difference in each student's actual performance on the statewide FSA assessments from that student's expected performance, which is supposed to account for student and classroom factors that impact the learning process. Aggregated value-added measures are created and averaged to the individual teacher in order to then create a teacher value-added measure (Basileo & Toth, 2019). Detractors say VAMs and LGMs reduce education to test performance, poison workplace relations by setting teachers in competition, and are opaque, biased, and unreliable in practice (Amrein-Beardsley, 2014). "The VAMs and LGMs have been controversial not only because the technology challenges the professional autonomy of teachers but also because experts such as economists, statisticians, and education policy specialists disagree over its scientific legitimacy" (Griffen & Panofsky, 2020, p. 445).

According to the Florida Student Growth Implementation Committee (FSGIC) Report (2015), in the VAM Equation Formula, school and teacher effects were treated as random effects, and the K-12 public school teacher and school-specific values are empirical Bayes estimates.

The FLDOE developed the model and began applying it during the 2011-12 school year. It is designed to measure a teacher's effectiveness. Most K-12 public school teachers in the state gets a VAM score. That score is based on the year-to-year performance of students on standardized tests, and on how those results reflect on teachers (FLDOE, 2020c). Griffen and Panofsky (2020) assert that in essence, a student's results largely determine a teacher's "value," regardless of that student's social, economic or psychological circumstances. "The formula is very controversial, because it relies on seemingly mathematical measures that are, in fact, premised on numerous arbitrary variables that may or may not have any real value in determining a teacher's effectiveness" (Griffen & Panofsky, 2020, p. 451). The state's handling of the formula has also been controversial. While the state-mandated VAM requirement was removed in 2017 in lieu of locally created LGMs, many districts are still using it because the state statute still requires that teachers be evaluated based on student progression and it is difficult to find a way to do this given the lack of resources provided by the State of Florida to meet these statutory requirements (United for Florida Children, 2021). Teacher VAM and LGM scores are not publicly released in Florida. However, the annual DSGs which are reported by all of the K-12 public schools within each district to the FLDOE are a commensurate approximation to a teacher's VAM/LGM score at each individual school site. Teachers who do not teach in the specified FSA subject areas categories of student assessment must rely on the school score instead, therefore DSGs serve as a strong indicator of that component domain of the overall teacher evaluation rating.

In the last five years, at least eighteen VAM-related lawsuits have been brought before state and federal courts in ten different states. Most of these lawsuits are in states

where VAM has become well institutionalized as a way of making high-stakes decisions about teachers regarding their employment and salary. It includes states such as California, Florida, Louisiana, Minnesota, Nevada, New Mexico, New York, North Carolina, Tennessee, and Texas. New cases keep emerging and similar cases in the same state are combined due to overlapping claims. “Though not all of these cases have been completed, they all represent contests over the legitimacy of VAMs and LGMs as a scientific object” (Silbey, 2008; Diamond & Lempert, 2018, p. 11).

Chapter Summary

The multi-layered elemental framework contained within this literature review forms the foundation for further investigative research regarding the comprehensive evaluative K-12 teacher instructional personnel rating assessment that is utilized in the State of Florida and how its components are linked to the outcomes of the merit-based pay model salary schemes that are presently being implemented. Each Florida K-12 public school district (i.e. Beta and Pi) provides a model for how the input variables of teacher instructional personnel evaluation ratings, FSA Student Assessments (R%H and M%H) and DSGs which are commensurate with an approximation to VAM/LGM scores are calculated to help formulate the output variables of teacher personnel final evaluation rating scores of Highly Effective and Effective. The input variables such as R%H, M%H, and DSGs were analyzed to determine its impact upon the output variables of Teacher Instructional Personnel Evaluation Category Rating Percentages in the domains of THER and TER. The aforementioned evaluation rating domains are the only two that allow for annual merit pay monetary compensation for teachers in the K-12 public school system. The DSGs were examined since the FLDOE

does not release official teacher VAM score data to the public. These tests and analyses are essential in order to determine how the current assessment and evaluation systems impact teacher placement on K-12 public education merit pay-for-performance salary schedules in the state.

CHAPTER 3. RESEARCH DESIGN AND METHODOLOGY

The purpose of this study was to analyze teacher pay across two similar K-12 public school districts within the State of Florida. The selected districts served as a reflective indicator of the teacher evaluation methodology utilized in K-12 public education throughout the entire state. This research study sought to determine if the districts maintained comparable student assessment scores in English in the domain of Reading % Satisfactory or Higher (R%H), Mathematics % Satisfactory or Higher (M%H), and District School Grades (DSGs) which represent commensurate VAM and LGM scores. Since the FLDOE does not release individual K-12 public school teacher VAM and LGM scores, the DSGs were the most appropriate comparative score to utilize when comparing these two districts. Furthermore, this study endeavored to determine if there were any significant differences of the means for each of the teacher rating outcome scores of Highly Effective (THER) and Effective (TER) between both districts that determine placement upon K-12 pay-for-performance salary schedules. These are also the primary variables utilized by the Florida K-12 Public School Accountability Programs (FSRP and the FDAP) that directly impact merit pay.

Quantitatively this means that this study analyzed if there was a statistically significant difference of the means between the two K-12 public school districts in the (x)-input variable domains of FSA data in R%H, M%H, and DSGs that are components of the Florida Accountability Programs (FSRP and the FDAP) that fund and impact merit pay salary schedules. Additionally, this research investigated if any statistically

significant percentage differences of the means between districts existed in the (*y*)-output variables of teacher instructional personnel evaluation ratings in the THER and TER category domains that serve as predictors of authentic student achievement in relation to merit pay incentivization. These are the only two rating categorical domains that qualify for placement out of the five-tier evaluation system for K-12 public school merit pay plans in the state. Moreover, the FLDOE and the districts do not publish individual teacher VAM/LGM scores. For this reason, DSGs were analyzed as they are commensurate with an approximation to individual teacher VAM/LGM scores. The input and output variables were examined and analyzed between two large urban based K-12 public school districts in Florida.

The study sought to answer the following empirical research questions:

1. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA English in the domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts?
2. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts?
3. Is there a statistically significant difference of the means between the (*x*)-input variable of the DSGs between two large urban based Florida K-12 Public School Districts?

4. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (*y*)-output variable of the Highly Effective Category Designation (THER) between two large urban based Florida K-12 Public School Districts?
5. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (*y*)-output variable of the Effective Category Designation (TER) between two large urban based Florida K-12 Public School Districts?

The five empirical research questions were addressable by a null hypothesis with the K-12 public schools from both districts serving as the unit of analysis. The district data are reported to the FLDOE. Accordingly, this research posited five null hypotheses:

$H_0$1. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA English in the domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts.

$H_0$2. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA Domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts.

$H_0$3. There is no statistically significant difference of the means between the (*x*)-input variable of the DSGs between two large urban based Florida K-12 Public School Districts.

$H_0$4. There is no statistically significant difference of the means between the (*y*)-output variable of the Highly Effective (THER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

$H_0$5. There is no statistically significant difference of the means between the (*y*)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

Site Sample

The sample for this study included all five hundred and one K-12 public schools in both districts that were used as the unit of analysis for the three (*x*)-input variables of R%H, M%H, and DSGs which are commensurate with an approximation to the VAM and LGM scores. The annually published district teacher instructional evaluation ratings by school that are reported to FLDOE were used as the unit of analysis for the (*y*)-output variables of the K-12 teacher instructional personnel evaluation ratings data (THER and TER) for two large urban based Florida K-12 public school districts (i.e. Beta and Pi). In Florida, the two large urban K-12 public school districts, Beta and Pi, were selected because they are ranked as two of the largest K-12 public school districts in the nation (FLDOE, 2021a), have comparable student demographics and utilize contrasting weighted merit pay salary schedule profiles and algorithms which could provide insights into the relationship between accountability and merit-based pay. Charter schools as well as public virtual schools were also included in this sample since they fall under the auspices of a K-12 state public school. This study was designed to determine through

independent samples t-tests if a statistically significant difference of the means between districts existed in the input variables of FSA R%H, M%H, and DSGs which are commensurate with an approximation to the VAM and LGM scores. The empirical data analyses of the output variables of THER and TER were conducted to determine if there was a significant difference in the percentage of teachers rated THER and TER between these large urban based K-12 public school districts. The input variables are a part of the Florida Accountability Programs (FSRP and the FDAP) which affect the outcome variables that determine placement in Merit-Based Pay Salary Schedules throughout the state. This analysis occurred over seven academic school years from 2012-2019. Each of the K-12 public school districts that was selected for this study participated in all years of the FSRP and FDAP Programs. The Independent Samples t-tests were conducted to address the research questions and the null hypotheses. Additionally, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were utilized to ascertain supplementary contextual information about the relationship between the tested input and output variables between and within both districts. The intercorrelation information regarding the input variables served to bolster the arguments made in relation to the output variable analyses.

Methods

Independent Sample t-tests, Comparative Means and Standard Deviation analyses were performed with all of the K-12 public schools within both districts employed as the unit of analysis. This was done to determine if there was a statistically significant difference of the means between districts in each of the respective input and output variable categories. Additionally, Intercorrelation Matrices, Independent Correlation

Contrasts, and Overlapping Dependent Correlation Contrasts were constructed for the input variables (R%H, M%H, and DSGs) and the output variables (THER and TER) to attain information regarding the relationships between the input and output variables within and between both districts. This was done to determine if there was a significant difference in the percentage of teachers rated THER and TER between both districts. The statistical tests run on the input variable data related specifically to the output variable data of the teacher instructional personnel ratings by school and were utilized as a result of their connection to the Florida Accountability Programs (FSRP and FDAP) that provide funding for the merit pay-for-performance salary schedules.

The tests were implemented to determine whether there was statistical evidence that would reveal if the accountability input variables (FSA R%H, M%H, and DSGs) means between districts and the output variable teacher ratings by school (Highly Effective and Effective) were significantly different. Moreover, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were constructed to provide ancillary information regarding the relationship between the input and output variables within and between both K-12 public school districts as it would determine placement upon merit pay salary schedules for K-12 public school teachers in Florida.

The data for all variables were originally maintained in a Microsoft Excel spreadsheet prior to being exported to the IBM SPSS Statistics (Version 28) predictive analytics software (2021) and Morris Statistical Software. The Alpha level was set at $\alpha = .01$ for all of the Independent Sample t-tests and the Intercorrelation Matrices in order to obtain a 99% confidence level in determining whether statistically significant

differences exist and for ancillary Intercorrelation Matrices information. The Independent Correlation Contrasts and Overlapping Dependent Correlation Contrasts were set at $\alpha = .05$ in regard to the input and output variable relationships. The FLDOE annually publishes accountability reports and statistical information about K-12 public education and employment outcomes for Florida students, teaching instructional personnel, and schools. These data are archived on the department's website and made accessible to the general public. Therefore, all pertinent data for this study were retrieved electronically and input into computer software programs for coding and analysis.

Quantitative methods and empirical data analyses included Independent Sample t-tests, means, and standard deviation inquiry. They were performed on the input variables (R%H, M%H, and DSGs) and the output variables (THER and TER) to respond to each of the five research questions and test the five corresponding null hypotheses between both K-12 public school districts. As an addendum, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were implemented to obtain ancillary information regarding the relationship between the input and output variables within and between both districts. The independent and dependent correlation contrasts addressed whether there were any statistically significant differences between the selected correlations that existed between the input variables of the Florida Accountability Programs (FSRP and the FDAP) and the outcome variables of the percentage of teachers rated Highly Effective and Effective between both K-12 public school districts. The research study was trying to determine the correlation of the input variables to the output variables within and between both K-12 public school districts. One component of the study examined the null hypotheses in regard to the differences in

means. The second component considered the overlapping dependent correlation contrasts because one variable was in common with both correlations. A total of five hundred and one schools in both Florida K-12 public school districts were identified for FSA (R%H, M%H) and DSGs data collection as the unit of analysis for the three input variables in conjunction with the output variables of the teacher instructional performance evaluation ratings of THER and TER. The annual DSGs which are reported by all of the K-12 public schools within each district to the FLDOE are commensurate with an approximation to teachers' VAM/LGM scores at each individual school site. This categorical domain was used because the FLDOE does not release teacher VAM/LGM scores to the public. During this seven-year longitudinal study, five independent variables were identified. The input variables were defined as R%H, M%H, and DSGs. The output variables were defined as the Highly Effective and Effective Teacher Performance Category Ratings. These are the only two rating categories that qualify for placement upon the merit pay-for-performance salary schedules. The tests and analyses served as a microcosm of what is currently occurring throughout the United States.

Data Collection

The data collection for the five variables were retrieved electronically from the FLDOE PK-20 Information Portal and the school accountability archival databases that maintain annual records regarding the input variables of FSA Student Assessment Data (R%H, M%H) and DSGs for all schools contained within a K-12 public school district within the state. The output variables of Teacher Instructional Personnel Evaluation Ratings by School also revealed all 5-tiers (Highly Effective [THER], Effective [TER], Needs Improvement [NI], 3-Years Developing [3YD] and Unsatisfactory [U]) that are

reported annually by district to the FLDOE. All of these data were available on the FLDOE website. No K-12 public school district in Florida has 100% of the students tested each academic school year. If any of the K-12 public school districts that were analyzed had missing student assessment or DSGs data, in SPSS, the discrete missing data were coded as "999" for that particular school during the specified academic school year. All of the input variables (R%H, M%H, and DSGs) that were tested were counted and evaluated as components of the FSRP and FDAP that specifically relate to merit pay. The outcome variables (THER and TER) were analyzed and are the only two rating categories utilized to determine placement and advancement upon Merit Pay Salary Schedules implemented for K-12 public school educators in Florida.

Data were collected for each of the tested variables using the following procedures:

FSA Student Assessment (Input Variables)

All K-12 public schools in Florida are required to test students at each academic level (i.e. primary, middle, secondary). The FSA R%H and M%H achievement test scores are directly tied to merit pay in the state. Therefore, the input variables of R%H and M%H were examined between two comparable large urban based K-12 public school districts to serve as a microcosm of what may be occurring throughout the United States.

District School Grades (Input Variable)

According to the Florida Grading System Act (2021), this act can be found in the Florida Statute § 1008.34. All K-12 public schools within a district receive an annual school grade designated as A, B, C, D or F. Under the Statute – Section 2 – School

Grades - Schools shall be graded using one of the following grades, defined according to rules of the State Board of Education (FLDOE, 2021e):

(a) "A" schools making excellent progress.

(b) "B" schools making above average progress.

(c) "C" schools making satisfactory progress.

(d) "D" schools making less than satisfactory progress.

(e) "F" schools failing to make adequate progress.

Under Section 5 which adheres to the Florida Grading System Act (2021), this act can be found in the Florida Statute §1008.34 – District Grades. A district grade must include measures of the district's progress in demonstrating Learning Gains of its highest-performing students. Individual teacher VAM/LGM scores are not published by FLDOE. Therefore, the District School Grade (DSG) for each school was utilized as a commensurate approximation to the VAM/LGM score. Not all teacher instructional personnel receive an individual VAM/LGM score due to the subject that he or she teaches. When that situation arises, the teacher is given the school score (i.e. DSG) for their particular school site. The DSG data are an input variable component that was used as a part of the overall comprehensive teacher instructional personnel evaluation score. Since this is connected to placement upon merit pay salary schedules for K-12 public school teachers within the State of Florida, it was essential to also analyze this data.

Teacher Instructional Personnel Evaluation Ratings Data (Output Variables)

In the State of Florida, all K-12 public schools are required to annually evaluate their instructional teaching staff. In the state, the Marzano Evaluation Tool iObservation ® is utilized to designate a 5-tier teacher rating as either – Innovating (4), Applying (3),

Developing (2), Beginning (1), Not Using (0) and Not Applicable (FLDOE, 2021e). The aforementioned numerical scores that are used in the IP domain in the overall teacher evaluation ratings are then converted into one of the output variable components that comprise a part of the state's categorical evaluation ratings for K-12 public school teachers which is comprised of a 5-tier designation system of Highly Effective (THER), Effective (TER), Needs Improvement (NI), 3-Years Developing (3YD) or Unsatisfactory (U). Only the two (y)-output variable rating domain designations of THER and TER allow a K-12 public school teacher to qualify for placement and advancement on merit pay salary schedules.

The FLDOE annually publishes the input variables of FSA Data (R%H and M%H) and DSGs that are commensurate with an approximation to VAM/LGM scores as well as the output variables reflected in the Teacher Instructional Personnel Evaluation Rating by School Data (THER and TER) for the analyzed K-12 public school districts. The input and output variables were collected over a 7-year period (2012-2019). It is the most current information available because the FLDOE Emergency Order No. 2020-EO-1 states that the Spring K-12 statewide assessment test administrations for the 2019-20 school year were canceled and accountability measures reliant on such data were not calculated for the 2019-20 school year (FLDOE, 2020c). This was a result of the worldwide pandemic of Covid-19.

The data collected focused on the input variables of K-12 public school FSA in R%H, M%H, and DSGs and the output variable domains of teacher instructional personnel evaluation ratings (i.e. Highly Effective [THER], Effective [TER], Needs Improvement [NI], 3-Years Developing [3YD], and Unsatisfactory [U]) for both districts.

The study considered five variables in these large urban based K-12 public school districts for comparative analyses and all data were obtained from the FLDOE.

Data Analysis

The study analyzed data collected over a seven-year period (2012-2019) for all Florida K-12 public schools that is annually published through district data for two of the largest urban based K-12 public school districts in the United States (i.e. Beta and Pi). Many of the K-12 public school districts throughout the nation utilize the same or a very similar teacher evaluation model of combining student assessment, VAM/LGM score and the teacher instructional evaluative observation component - IP (i.e. Marzano's iObservation ® score) in association with merit pay-for-performance salary schemes. The study was designed to determine if there were statistically significant differences in the means of each of the input variables (R%H, M%H, and DSGs) and output variables (THER and TER) through Independent Sample t-tests between districts. The ancillary Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts addressed the correlations and contrasts between the input variables and each teacher rating output variable within and between both urban K-12 public school districts.

Once all of the data were collected for the tested input and output variables, the analyses were performed in the following manner:

FSA Student Assessment (Input Variables)

Independent Samples t-tests were conducted to address the research questions and null hypotheses and to determine if there were statistically significant differences of the means in the input variable domains of FSA R%H and M%H between the two large

comparable urban K-12 public school districts (i.e. Beta and Pi) for each academic school year from 2012-2019. Additionally, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were constructed as an addendum to ascertain information about the relationship between the input variables that provided information regarding the correlations and contrasts between student assessments and DSGs and each of the teacher rating output variables. The correlations were analyzed for each district separately and between the two combined districts. The data were analyzed with IBM SPSS Statistics (Version 28) predictive analytics software (2021) [Tables 1-8] and Morris Statistical Software (Morris, 2022) [Tables 9-14] in order to answer the research questions and null hypotheses 1, 2 and $H_0$1, $H_0$2.

District School Grades (Input Variable)

An Independent Samples t-test was also conducted to address this research question and null hypothesis and to determine if there were statistically significant differences of the means in the input variable domain of DSGs between the two large urban K-12 public school districts (i.e. Beta and Pi) for each academic school year from 2012-2019. Additionally, an Intercorrelation Matrix was constructed as an addendum to ascertain information about the relationship between the input variables which calculated the correlations for each district separately and for the two combined districts. Each of the two large urban K-12 public school districts (i.e. Beta and Pi) obtained a grade designation by the state. A numerical value was ascribed for each letter grade during the 7 years of study as follows: A = 4, B = 3, C = 2, D = 1 and F = 0. The data were analyzed with IBM SPSS Statistics (Version 28) predictive analytics software (Tables 1-8) and

Morris Statistical Software (Morris, 2022) [Tables 9-14] in order to answer the research questions and null hypotheses 3 and $H_0$3.

Teacher Instructional Personnel Evaluation Ratings Data (Output Variables)

Independent Samples t-tests, Comparative Means, and Standard Deviation analyses were conducted to address the research questions and null hypotheses and to determine if there were statistically significant differences of the means in the output variable domains of THER and TER between the two large comparable urban K-12 public school districts (i.e. Beta and Pi) for each academic school year from 2012-2019. Additionally, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were constructed as an addendum to ascertain information about the relationship between the output variables within and between both districts. The data were analyzed with IBM SPSS Statistics (Version 28) predictive analytics software (2021) [Tables 1-8] and Morris Statistical Software (Morris, 2022) [Tables 9-14] in order to answer the research questions and null hypotheses 4, 5 and $H_0$4, $H_0$5.

Chapter Summary

This chapter outlined and discussed the procedures that were used to collect and analyze 7-years of data for two large urban based Florida K-12 public school districts (i.e. Beta and Pi) that was included in this study. The research design employed "a quantitative and parametric methodology" (Kothari & Garg, 2019, p. 3) to determine if any statistical differences of the means between districts existed in the input variables (R%H, M%H, and DSGs) and the output variables (THER and TER). Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation

Contrasts were also constructed as an addendum to provide any ancillary information regarding the relationship between the input and output variables within and between both districts to determine if any significant difference existed in the percentage of K-12 public school teachers who were rated Highly Effective and Effective in relation to placement upon merit pay salary schedules between both districts. Chapter 4 revealed and delineated the findings of the statistical analyses.

CHAPTER 4. FINDINGS OF THE STUDY

The following section will discuss the findings of the input variables of English in the domain of Reading % Satisfactory or Higher (R%H), Mathematics % Satisfactory or Higher (M%H), and District School Grades (DGSs) that are utilized to formulate compensation packages based upon the output variables of teachers being classified as Highly Effective (THER) and Effective (TER) upon pay-for-performance merit pay salary schedules.

Descriptive Statistics

For this study, the K-12 public schools (Beta – 294 and Pi – 207; 501 in total) contained within each district were selected as the unit of analysis for the (x)-input variables (R%H, M%H, and DSGs) and the (y)-output variables (THER and TER). The numbers and percentages from the 5-tier instructional teacher evaluation rating system by school was reported annually by district to the FLDOE. In Florida, the two large urban K-12 public school districts, Beta and Pi, were selected because they are ranked as two of the largest school districts in the nation (FLDOE, 2021a), have comparable student demographics and utilize contrasting weighted merit pay salary schedule profiles and algorithms which could provide insights into the relationship between accountability and merit-based pay. Both K-12 public school districts had at least 98% of the student data reported annually for all variables by the FLDOE. Thus, as shown in Table 1 and Table 2, an examination of the minimum, maximum, mean, and standard deviation of the tested variables was performed.

Table 1

Descriptive Statistics of the Tested (x)-Input Variables Percentage Rates for (R%H, M%H and DSGs) Between Two K-12 Public Schools District Analysis Over a 7-Year Study

Input Variables	Minimum	Maximum	Mean	Standard Deviation
FSA English in the Domain of Reading % Satisfactory or Higher (R%H) Beta District (1) Pi District (2) 2012-2019	1) 55 2) 55	1) 60 2) 59	1) 57.43 2) 57.29 Mean Dif. Bet.District (.14)	1) 1.90 2) 1.89
FSA Math % Satisfactory or Higher (M%H) Beta District (1) Pi District (2) 2012-2019	1) 55 2) 56	1) 61 2) 63	1) 59.00 2) 60.57 Mean Dif. Bet.District (1.57)	1) 2.52 2) 2.70
District School Grades (DSGs) Beta District (1) Pi District (2) 2012-2019	1) 2 2) 3	1) 3 2) 4	1) 2.86 2) 3.43 Mean Dif. Bet.District (.57)	1) .38 2) .54

FSA English domain of Reading % Satisfactory or Higher [Input Variable]

The input variable of FSA R%H had a Beta District minimum of 55% and a maximum of 60% and a Pi District minimum of 55% and a maximum of 59%. The Beta District mean was 57.43% and the Pi District mean was 57.29%. The difference of the means between both K-12 school districts was (.14%). The standard deviation for the Beta District was 1.90% and the Pi District was 1.89%. This variable is utilized in the student assessment domain for merit pay calculations. The standard deviation means

difference was (.01%) between both K-12 public school districts in the input variable of R%H.

FSA Mathematics % Satisfactory or Higher [Input Variable]

The input variable of FSA M%H had a Beta District minimum of 55% and a maximum of 61%. The Pi District minimum was 56% and the maximum was 63%. The Beta District mean was 59% and the Pi mean was 60.57%. The difference of the means between both K-12 school districts was (1.57%). The standard deviation for the Beta District was 2.52% and the Pi District was 2.70% for both K-12 public school districts.

District School Grades [Input Variable]

The input variable of DSGs (utilized as a VAM/LGM commensurate assessment) was examined for a 7-year duration. It was coded as follows: A = 4, B =3, C = 2, D = 1, F = 0. The Beta District mean was 2.86 and the minimum and maximum ranged from 2 to 3. The Pi District mean was 3.43 and the minimum and maximum ranged from 3 to 4. The difference of the means between both K-12 school districts was (.57). The standard deviation for the Beta District was (.38) and the Pi District was (.54). The mean and standard deviation for this input variable was calculated from data provided by the FLDOE for each K-12 public school district. The DSG served as a commensurate approximation to the VAM/LGM score component that is factored into the teaching instructional personnel overall evaluation rating data score since individual teacher VAM/LGM scores are not reported to the public. This input variable component was also utilized to determine merit pay calculations for all K-12 public school teachers in both large urban based Florida K-12 public school districts.

Table 2

Descriptive Statistics of the Tested (y)-Output Variables in the Number of Teacher Instructional Personnel Evaluation Ratings That Qualify for Merit Pay Between Two K-12 Public School Districts Over a 7-Year Study

Output Variables	Minimum	Maximum	Mean	Standard Deviation
Teacher Instructional Rating – (THER) Highly Effective Beta District (1) Pi District (2) 2012-2019	1) 736 2) 4.4K	1) 10.0K 2) 11.5K	1) 3.6K 2) 6.7K	1) 3.2K 2) 2.5K
Teacher Instructional Rating – (TER) Effective Beta District (1) Pi District (2) 2012-2019	1) 5.7K 2) 516	1) 13.7K 2) 7.1K	1) 11.4K 2) 5.0K	1) 2.7K 2) 2.3K

Teacher Highly Effective Rating [Output Variable]

The unit of analysis for this output variable were the annually reported district teacher instructional personnel evaluation ratings by school to the FLDOE. The output variable of THER which represented the teaching instructional faculty personnel from both of the large urban K-12 public school districts (i.e. Beta and Pi) presented a Beta District mean of 3,569 and a Pi District mean of 6,658 representing the average of the total number of teachers that were rated Highly Effective over a 7-year period in each district (FLDOE, 2021b; Pi District Schools Budget Department, 2021). By examining each district separately over the same 7-year period, Beta rated 24,981 teachers in total as Highly Effective over a 7-year period and Pi rated 46,606 as Highly Effective over the

same 7-year time duration. During this 7-year longitudinal study, the difference between both large urban K-12 public school districts in Florida is that Pi rated 21,625 more teachers as Highly Effective in comparison to Beta.

Teacher Effective Rating [Output Variable]

The unit of analysis for this output variable was the annually reported district teacher instructional personnel evaluation ratings by school data to the FLDOE. The output variable of TER which represented the teaching instructional faculty personnel from both of the K-12 public school districts (i.e. Beta and Pi) presented a Beta District mean of 11,448 and a Pi District mean of 5,031 representing the average of the total number of teachers rated Effective over a 7-year period. By examining each district separately over the same 7-year period, Beta rated 80,119 teachers in total as Effective over a 7-year period and Pi rated 35,214 as Effective over the same 7-year time duration. During this 7-year longitudinal study, the difference between both large urban K-12 public school districts in Florida is that Beta rated 44,905 more teachers as Effective in comparison to Pi (BSD, 2021a; FLDOE, 2021b).

Each of the output variables, THER and TER, determine placement and advancement upon the merit pay salary schedules in each of the respective K-12 public school districts that were analyzed. As shown in Table 7 and Table 8, the data were presented to the FLDOE by each of the tested districts in relation to the teacher evaluation rating data and the primary variables that were used to calculate those output variable category designations.

Table 3

Teacher Instructional Personnel Evaluation Rating Data (5-Tier System) by Florida District

Teacher Evaluation Data										
	Highly Effective	Effective	Needs Improvement	3 Years Developing	Unsatisfactory	Not Evaluated	Total Teacher Count	Reading % Satisfactory or Higher	Math % Satisfactory or Higher	District Grade
Beta										
2012-2013	9.9%	89.1%	0.8%	0.2%	0.0%	13.5%		58	61	C
	1,526	13,765	116	34	4	2,091	17,536			
2013-2014	5.3%	93.9%	0.5%	0.1%	0.2%	16.2%		59	61	B
	736	13,158	67	20	25	2702	16,708			
2014-2015	14%	85.1%	0.3%	0.5%	0.1%	14.1%		55	55	B
	2,040	12,403	50	69	12	2,399	16,973			
2015-2016	18.3%	80.3%	1.3%	0.0%	0.1%	17.2%		55	56	B
	2,656	11,645	188	5	8	3,005	17,507			
2016-2017	19.6%	78.9%	0.9%	0.4%	0.2%	13.8%		57	59	B
	3,018	12,159	132	58	37	2,457	17,861			
2017-2018	29.2%	70.1%	0.5%	0.1%	0.1%	8.7%		58	60	B
	4,725	11,332	81	24	10	1,534	17,706			
2018-2019	64.1%	35.3%	0.5%	0.1%	0.1%	0.5%		60	61	B
	10,280	5,657	73	18	13	81	16,122			
Pi										
2012-2013	37.9%	61.8%	0.0%	0.2%	0.0%	10%		59	62	B
	4,350	7,090	4	18	5	1,272	12,739			
2013-2014	43.6%	56.2%	0.0%	0.2%	0.0%	11.1%		59	63	B
	4,964	6,392	0	22	1	1,414	12,793			
2014-2015	41.9%	57.3%	0.4%	0.4%	0.0%	12.2%		55	56	A
	4,828	6,592	41	50	1	1,605	13,117			
2015-2016	50.4%	49.1%	0.2%	0.4	0.0%	10.8%		55	58	B
	5,934	5,779	18	43	1	1,428	13,203			
2016-2017	57.2%	42.0%	0.1%	0.6%	0.1%	9.6%		56	60	B
	6,845	5,026	17	69	12	1,273	13,242			
2017-2018	67.6%	31.6%	0.2%	0.5%	0.1%	4.7%		58	62	A
	8,155	3,819	28	58	9	594	12,663			
2018-2019	95.5%	4.3%	0.1%	0.1%	0.1%	1.2%		59	63	A
	11,530	516	13	10	7	150	12,226			

Note. Table created in SPSS from selected district data retrieved electronically from the FLDOE. Copyright 2012-2019.

This quantitative study determined through Independent Samples t-tests that there were no statistically significant differences on the means of each of the (*x*)-input variables (R%H, M%H, and DSGs) between districts. However, there were statistically significant differences on the means of each of the (*y*)-output variables (THER and TER). Additionally, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were constructed as an addendum to

ascertain information about the relationship between the input and output variables. The data showed that the input variables of R%H, M%H, and DSGs were correlated within each individual district as well as districts combined. However, there were differences of the means in the percentage of K-12 teachers rated in the (*y*)-output variables (THER and TER) in both of the K-12 public school districts' respective variable domain comparisons. The input variables served as components of the Florida Accountability Programs (FSRP and the FDAP) that provides financially incentivized support that impacts merit pay salary schedules throughout the state.

In this seven-year longitudinal analysis, the Independent Samples t-tests revealed that the output variables between both K-12 public school districts displayed significant differences in the percentage of K-12 public school teachers who are rated THER and TER. The Independent Correlation Contrasts and the Overlapping Dependent Contrasts provided important information regarding the correlations and contrasts between the input variables and the output variables in their respective teacher evaluation rating categories. Although the following is not a result of the testing and analyses, it is essential to understand the context for the selection of the analyzed districts. In Florida, the two large urban K-12 public school districts, Beta and Pi, were be selected because they are ranked as two of the largest school districts in the nation (FLDOE, 2021a), have comparable student demographics and utilize contrasting weighted merit pay salary schedule profiles and algorithms which could provide insights into the relationship between accountability and merit-based pay.

The Independent Samples t-tests revealed that there were no statistically significant differences of the means between the selected input variables (R%H, M%H,

and DSGs) of the FSRP and FDAP for each K-12 public school district that would impact components of the output variables of teacher evaluation ratings (THER and TER) which serve to determine placement of K-12 public school teachers on merit pay-for-performance salary schedules. The output variable outcomes were not consistent with the input variable data. In this study, a total of 501 (Beta – 294 and Pi – 207) schools were the unit of analysis for each of the input and output variables. The annual district numbers and percentages of the teacher instructional personnel evaluation ratings by school was provided to the FLDOE. Over this seven-year study, the FSRP and FDAP input variable categories (R%H, M%H, DSGs) and the output variables ([THER] and [TER] were tested and analyzed.

Hypothesis Testing

Through the utilization of the IBM SPSS Statistics (Version 28) predictive analytics software, two statistical tests and multiple analysis methods were employed to analyze the collected data. These analyses included Independent Samples t-tests for all of the input variables (R%H, M%H, and DSGs) and output variables (THER and TER) which produced a means and standard deviation comparison. Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts served as an addendum to ascertain information about the relationship between the input and output variables within each district and by a correlational comparison between districts. Tests and analyses were performed on the input (R%H, M%H, and DSGs) and the output (THER and TER) variables obtained from data from the FLDOE that are utilized directly in merit pay salary schedule placement.

Research Questions and Null Hypotheses

Each of the five empirical research questions were addressable by a null hypothesis with the K-12 public schools from both districts serving as the unit of analysis that was reported to the FLDOE. Therefore, this research posited five null hypotheses. The research questions and corresponding null hypotheses were as follows:

1. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA Domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts?

$H_0$1. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA Domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts.

An Independent Samples t-test was conducted to compare the input variable of FSA R%H data between two K-12 public school districts. The Table 4 revealed that the Levene's Test confirmed the Homogeneity of Variance (HOV) for this input variable. Over the 7-year longitudinal study (2012-2019), there was not a significant difference of the means of the input variable, R%H, between both K-12 public school districts. The Cohen's *d* indicated a very small effect size.

Table 4

Independent Samples t-test Performed on the (x)-Input Variable Category of English in the Domain of Reading % Satisfactory or Higher (R%H) Between Two K-12 Public School Districts Over a 7-Year Study (2012-2019)

Input Variable District Code	Levene's HOV					
	F	*p*	*t*	*df*	*p*	Cohen's *d*
Reading % Satisfactory or Higher (R%H) 2012-2019	4.74	.12	.05	492	.52	<.01

Note. 99% K-12 student population tested in this study. Values reflect the average across 7 years of data.
** p < .01

2. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA Domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts?

$H_0$2. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA Domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts.

An Independent Samples t-test was conducted to compare the input variable of FSA M%H data between two K-12 public school districts. The Table 5 revealed that the Levene's Test confirms the Homogeneity of Variance (HOV) for this input variable.

Over the 7-year longitudinal study (2012-2019), there was not a significant difference of the means of the input variable, M%H, between both K-12 public school districts. The Cohen's *d* indicated a very small effect size.

Table 5

Independent Samples t-test Performed on the (x)-Input Variable Category of Mathematics % Satisfactory or Higher (M%H) Between Two K-12 Public School Districts Over a 7-Year Study (2012-2019)

Input Variable District Code	Levene's HOV					
	F	*p*	*t*	*df*	*p*	Cohen's *d*
Mathematics % Satisfactory or Higher (M%H) 2012-2019	8.83	.38	.87	491	.38	.08

Note. 98% K-12 student population tested in this study. Values reflect the average across 7 years of data.
** p < .01

3. Is there a statistically significant difference of the means between the (*x*)-input variable of the DSGs between two large urban based Florida K-12 Public School Districts?

$H_0$3. There is no statistically significant difference of the means between the (*x*)-input variable of the DSGs between two large urban based Florida K-12 Public School Districts.

An Independent Samples t-test was conducted to compare the input variable of DSGs data between two K-12 public school districts. The Table 6 revealed that the Levene's Test confirms the Homogeneity of Variance (HOV) for this input variable. Over the 7-year longitudinal study (2012-2019), there was not a significant difference of the means of the input variable, DSGs, between both K-12 public school districts. The Cohen's *d* indicated a very small effect size.

Table 6

Independent Samples t-test Performed on the (x)-Input Variable Category of District School Grades (DSGs) Between Two K-12 Public School Districts Over a 7-Year Study (2012-2019)

Input Variable District Code	Levene's HOV					
	F	*p*	*t*	*df*	*p*	Cohen's *d*
District School Grades (DSGs) 2012-2019	1.10	.41	1.23	471	.16	.12

Note. 98% K-12 student population tested in this study. Values reflect the average across 7 years of data.
** p < .01

4. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (*y*)-output variable of the Highly Effective (THER) Category Rating Designation between two large urban based Florida K-12 Public School Districts?

$H_0$4. There is no statistically significant difference in the percentage of the

mean between the (*y*)-output variable of the Highly Effective (THER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

An Independent Samples t-test was conducted to compare the output variable of the THER data of all Instructional Teachers between two K-12 public school districts. Table 7 revealed that the Levene's Test in relation to the Homogeneity of Variance (HOV) that due to Welch, the Equal Variances was not assumed for this output variable. The Cohen's *d* indicated that the magnitude of effect was large. Therefore, the means were significantly different with a large effect size. Over the 7-year longitudinal study (2012-2019), the significance level (*p*) for the Independent Samples t-test was significant. The direction of the significant effect showed that the Pi District had the higher mean and the effect size was large. Consequently, it could be argued that there is a statistically significant difference of the means in the percentage between the (*y*)-output variable of THER Category Rating Designation between two large urban based Florida K-12 Public School Districts. The two distributions of teacher evaluation ratings in the domain of THER were significantly different.

Table 7

Independent Samples t-test Performed on the (y)-Output Variable Category of Highly Effective (THER) Teacher Rating Between Two K-12 Public School Districts Over a 7-Year Study (2012-2019)

Output Variable District Code	Levene's HOV					
	F	*p*	*t*	*df*	*p*	Cohen's *d*
Highly Effective Teacher Rating (THER) 2012-2019	115.99	.02	-8.13	382	<.01**	.79

Note. 99% K-12 student population tested in this study. Values reflect the average across 7 years of data.
** p < .01

5. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (*y*)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts?

$H_0$5. There is no statistically significant difference in the percentage of the means between the (*y*)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

An Independent Samples t-test was conducted to compare the output variable of the Effective Teacher Rating (TER) data of all Instructional Teachers between two

K-12 public school districts. Table 8 revealed that the Levene's Test in relation to the Homogeneity of Variance (HOV) that due to Welch, the Equal Variances was not assumed for this output variable. The Cohen's *d* indicated that the magnitude of effect was medium. Over the 7-year longitudinal study (2012-2019), the significance level (*p*) for the Independent Samples t-test was significant. The direction of the significant effect showed that the Beta District had the higher mean and the effect size was medium. Consequently, it could be argued that there is a statistically significant difference of the means in the percentage between the (*y*)-output variable of Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts. The two distributions of teacher evaluation ratings in the domain of Effective (TER) were significantly different.

Table 8

Independent Samples t-test Performed on the (y)-Output Variable Category of Effective (TER) Teacher Rating Between Two K-12 Public School Districts Over a 7-Year Study (2012-2019)

Output Variable District Code	Levene's HOV					
	F	*p*	*t*	*df*	*p*	Cohen's *d*
Effective Teacher Rating (TER) 2012-2019	29.63	.16	5.59	529	<.01**	.66

Note. 99% K-12 student population tested in this study. Values reflect the average across 7 years of data.
** p < .01

Table 9 revealed that FSA R%H had positive correlation with FSA M%H that was significant. DSGs and the THER Teacher Rating were negatively correlated and was not significant to the FSA (R%H and M%H). However, the TER was positively correlated to FSA R%H, FSA M%H, and DSGs but was not significant within the Beta District.

Table 9

Descriptive Statistics and Bivariate Intercorrelation Matrix Test Performed on the (x)- Input Variables (R%H, M%H, and DSGs) and the (y)-Output Variables (THER and TER) in the Beta K-12 Public School District Over a 7-Year Study (2012-2019)

Beta Input Variables (R%H, M%H, and DSGs) Beta Output Variables (THER and TER)	*M*	*SD*	1	2	3	4	5
1. English in the domain of Reading % Satisfactory or Higher (R%H) 2012-2019	57.43	1.90	_				
2. Mathematics % Satisfactory or Higher (M%H) 2012-2019	59.00	2.52	.94**	_			
3. District School Grades (DSGs) 2012-2019	2.86	.38	-.13	-.35	_		

Beta Input Variables (R%H, M%H, and DSGs) Beta Output Variables (THER and TER)	*M*	*SD*	1	2	3	4	5
4. Highly Effective Teacher Evaluation Rating (THER) 2012-2019	3,569	3,216	-.48	-.26	-.29	_	
5. Effective Teacher Evaluation Rating (TER) 2012-2019	11,448	2,687	.47	.25	.29	-.23**	_

Note. N = 294. 98% K-12 student population tested in this study. Values reflect the average across 7 years of data. ** $p < .01$, two-tailed.

Table 10 revealed that FSA M%H had positive correlation with FSA R%H and was significant. DSGs and the TER Rating Domain were negatively correlated and was not significant to the FSA R%H and M%H. However, the THER Rating Domain was positively correlated to FSA R%H, FSA M%H, and DSGs but was not significant within the Pi District.

Table 10

Descriptive Statistics and Bivariate Intercorrelation Matrix Test Performed on the (x)-Input Variables (R%H, M%H, and DSGs) and the (y)-Output Variables (THER and TER) in the Pi K-12 Public School District Over a 7-Year Study (2012-2019)

Pi Input Variables (R%H, M%H, and DSGs) Beta Output Variables (THER and TER)	*M*	*SD*	1	2	3	4	5
1. English in the domain of Reading % Satisfactory or Higher (R%H) 2012-2019	57.29	1.89	_				
2. Mathematics % Satisfactory or Higher (M%H) 2012-2019	60.57	2.70	.42**	_			
3. District School Grades (DSGs) 2012-2019	3.43	.54	-.02	-.34	_		
4. Highly Effective Teacher Evaluation Rating (THER) 2012-2019	6,658	2,519	.30	.31	.56	_	
5. Effective Teacher Evaluation Rating (TER) 2012-2019	5,031	2,271	-.28	-.33	-.55	-.02**	_

Note. N = 207. 98% K-12 student population tested in this study. Values reflect the average across 7 years of data. ** $p < .01$, two-tailed.

Table 11 revealed that FSA R%H in the combination of both the Beta and Pi K-12 public school districts had significant positive correlation with FSA M%H in both districts. The Beta and Pi District DSGs were negatively correlated and were not

significant to any FSA (R%H and M%H) as well as to each other in both districts. The Pearson Correlation between the THER Teacher Rating Designation was also positively correlated while the TER Teacher Rating was negatively correlated to FSA R%H, FSA M%H, and DSGs between the Beta and Pi Districts.

Table 11

Bivariate Intercorrelation Matrix Test Performed on the (x)-Input Variables (R%H, M%H, and DSGs) and the (y)-Output Variables (THER and TER) in Two K-12 Public School Districts Over a 7-Year Study (2012-2019)

Input Variables (R%H, M%H, and DSGs) Output Variables (THER and TER)	*M*	*SD*	1	2	3	4	5
1. English in the domain of Reading % Satisfactory or Higher (R%H) 2012-2019	57.36	1.90	_				
2. Mathematics % Satisfactory or Higher (M%H) 2012-2019	59.79	2.61	.54*	_			
3. District School Grades (DSGs) 2012-2019	3.15	.46	-.06	-.03	_		
4. Highly Effective Teacher Evaluation Rating (THER) 2012-2019	5,114	2,868	.11	.14	.09	_	
5. Effective Teacher Evaluation Rating (TER) 2012-2019	8,240	2,479	-.25	-.48	-.65*	-.33	_

Note. N = 489. 98% K-12 student population tested in this study. Values reflect the average across 7 years of data. * $p < .05$, two-tailed.

Fisher's (1921) test of the difference between independent correlations for input and output variables is included in Table 12. For the purposes of this research, Table 12 was constructed to determine whether the correlations in the districts are significantly different between the input (R%H, M%H, and DSGs) and the output variables (THER and TER). Therefore, this analysis sought to determine if the correlations between the input variables (English, Math, and DSGs) and the Output 1 (Highly Effective) and Output 2 (Effective) variables were significantly different in the two K-12 public school districts. Confidence Intervals (*CI*) were created using the method of Zou (2007) for each contrast. The null hypotheses had an $\alpha = .05$ and the *CI*s were set at a 95% level for the data contained in each of the following tables (i.e. Table 12). Independent Correlation Contrasts between the two K-12 Public School Districts for the (*x*)-Input Variables (R%H, M%H, and DSGs) and the (*y*)-Output Variables (THER and TER) were performed utilizing Morris Statistical Software (2022).

Table 12 revealed that the (*p*) for all differences in correlations were significantly different. The data showed the direction of the effect as the correlation in the Pi District was significantly larger than in the Beta District. The nature of these correlations was that English and Math were more strongly related in the Pi District than in the Beta District.

Table 12

Independent Contrasts Between the Two K-12 Public School Districts on Correlations Between the (x)-Input Variables (R%H, M%H, and DSGs) and the (y)-Output Variables (THER and TER) Over a 7-Year Study (2012-2019)

Input Variables (R%H, M%H, and DSGs) Output Variables (THER and TER)	*z*	*p*	$r_{(Beta)} - r_{(Pi)}$	*CI*
1. English (R%H) To Output 1 – Highly Effective (THER) 2012-2019	-9.12	<.0001*	-.78	[-.93, -.62]
2. Mathematics (M%H) To Output 1 – Highly Effective (THER) 2012-2019	-6.43	<.0001*	-.57	[-.73, -.40]
3. District School Grades (DSGs) To Output 1 – Highly Effective (THER) 2012-2019	-10.20	<.0001*	-.85	[-.98, -.70]
4. English (R%H) To Output 2 – Effective (TER) 2012-2019	8.74	<.0001*	.75	[.59, .90]
5. Math (M%H) To Output 2 – Effective (TER) 2012-2019	6.55	<.0001*	.58	[.41, .74]
6. District School Grades (DSGs) To Output 2 – Effective (TER) 2012-2019	10.04	<.0001*	.84	[.69, .98]

Note. N = 501. Values reflect the average across 7 years of data. * p < .05, two-tailed.

According to Raghunathan et al. (1996), comparisons between correlations are considered overlapping when in each comparison, "one of the two variables being

correlated with the other is also involved in the other correlation" (p. 178). In the following Tables 13 and 14, the overlapping dependent correlation contrasts were used to investigate which of the input variables (i.e. R%H, M%H, and DSGs) were better predictors of the output teacher rating variables (i.e. THER and TER) for each district. In order to obtain a more comprehensive, relevant, and holistic perspective regarding each of the analyzed districts, it was necessary to run the overlapping dependent correlation contrasts in this manner for the purpose of clarity to discern what occurred in each district in regard to the relationships between the input and output variables.

Zou (2007) suggested that Confidence Intervals (*CI*) as well as a comparison of the "*r*" encompass significance tests and provide an estimate of the magnitude of the effect (p. 399). Moreover, Zou (2007) came up with mechanisms for estimating the *CI*s for the difference in correlations in all correlation contrast situations (i.e. Independent Correlation Contrasts and Overlapping Dependent Correlation Contrasts). In terms of the two Overlapping Dependent Correlation Contrasts (Tables 13-14), Meng et al. (1992) maintain that it is also important "to provide simple but accurate methods for comparing correlation coefficients between a dependent variable and a set of independent variables" (p. 172). This research endeavored to determine whether one of the input variables (English or Math) was a stronger predictor of the output teacher evaluation ratings of Outcome Overlapping Dependent Variable 1 (Highly Effective) and Outcome Overlapping Dependent Variable 2 (Effective) respectively within districts or combined.

For the Overlapping Dependent Correlation Contrasts, Table 13 revealed the difference in predictive validity in the Beta District. The only two overlapping dependent correlation contrasts that were not significant were for input variables M%H and DSGs to

the variable for the teacher evaluation rating of Highly Effective and the variable signifying the teacher evaluation rating of Effective. Though the following difference in correlations was significant, the significance for the differences in correlations showed the direction of the effect as the correlation in the Beta District was significantly large. This research study compared the correlations between each input variable (R%H, M%H, and DSGs) and each output variable (THER and TER) solely within the Beta District. The nature of the difference in these correlations showed that English and Math had a significance that was more strongly related to both of the output variables (Highly Effective and Effective) in the Beta District.

Table 13

Overlapping Dependent Contrast Performed Between the (x)-Input Variables (R%H, M%H, and DSGs) and the (y)-Output Variables (THER and TER) by Individual K-12 Public School District (Beta District) Over a 7-Year Study (2012-2019)

Input Variables (R%H, M%H, and DSGs) Output Variables (THER and TER)	*z*	*p*	$r_{13} - r_{23}$	*CI*
1. English vs. Math In Relation To Output 1 Highly Effective	-11.70	<.0001*	-.22	[-.26, -.18]
2. English vs. District School Grades (DSGs) In Relation To Output 1 Highly Effective	-2.47	.01*	-.19	[-.34, -.04]

Input Variables (R%H, M%H, and DSGs) Output Variables (THER and TER)	*z*	*p*	$r_{13} - r_{23}$	*CI*
3. Math vs. District School Grades (DSGs) In Relation To Output 1 Highly Effective	.33	.74	.03	[-.15, .21]
4. English vs. Math In Relation To Output 2 Effective	11.65	<.0001*	.22	[.18, .26]
5. English vs. District School Grades (DSGs) In Relation To Output 2 Effective	2.33	.02*	.18	[.03, .33]
6. Math vs. District School Grades (DSGs) In Relation To Output 2 Effective	-.44	.66	-.04	[-.22, .14]

Note. N = 294. Values reflect the average across 7 years of data. * p < .05, two-tailed.

For the Overlapping Dependent Correlation Contrasts, Table 14 revealed the difference in predictive validity in the Pi District. The only two overlapping dependent contrasts that were not significant were for input variables R%H and M%H to the teacher evaluation rating of Highly Effective and the variable signifying the teacher evaluation rating of Effective respectively. Though the following correlations were significant, the significance for the differences in correlations showed the direction of the effect as the correlation in the Pi District was significantly large. This research study compared the difference in correlations between each input variable (R%H, M%H, and DSGs) and each output variable (THER and TER) solely within the Pi District. The nature of the difference in these correlations was that English and DSGs had a stronger significant

value to both of the output variables that reflected the teacher evaluation ratings (THER and TER) within in the Pi District.

Table 14

Overlapping Dependent Contrasts Performed Between the (x)-Input Variables (R%H, M%H, and DSGs) and the (y)-Output Variables (THER and TER) by Individual K-12 Public School District (Pi District) Over a 7-Year Study (2012-2019)

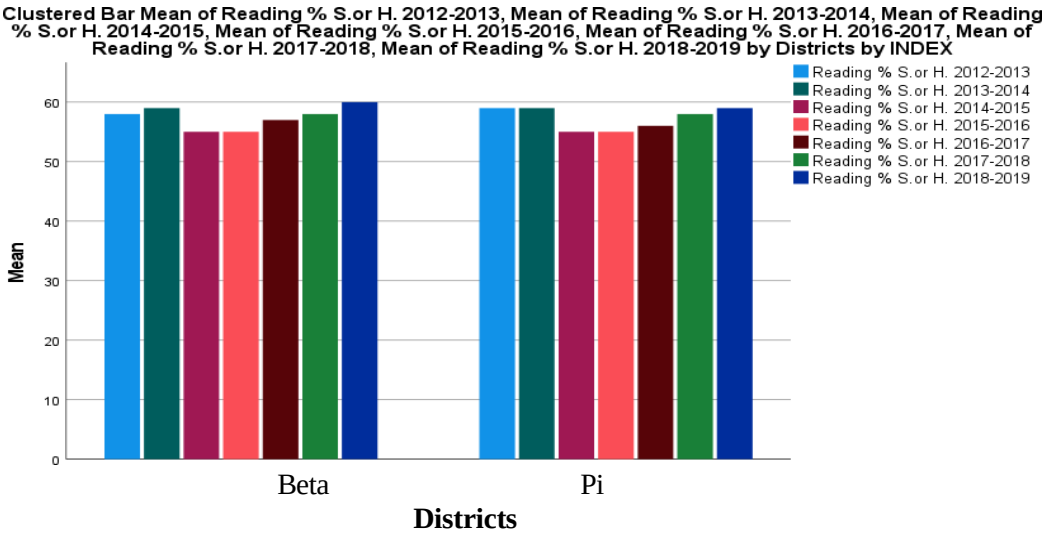
Input Variables (R%H, M%H and DSGs) Output Variables (THER and TER)	*z*	*p*	$r_{13} - r_{23}$	*CI*
1. English vs. Math In Relation To Output 1 Highly Effective	-.14	.89	-.01	[-.15, .13]
2. English vs. District School Grades (DSGs) In Relation To Output 1 Highly Effective	-3.10	<.01*	-.26	[-.42, -.10]
3. Math vs. District School Grades (DSGs) In Relation To Output 1 Highly Effective	-2.67	.01*	-.25	[-.43, -.07]
4. English vs. Math In Relation To Output 2 Effective	.71	.48	.05	[-.09, .19]
5. English vs. District School Grades (DSGs) In Relation To Output 2 Effective	3.17	<.01*	.27	[.10, .44]
6. Math vs. District School Grades (DSGs) In Relation To Output 2 Effective	2.36	.02*	.22	[.04, .40]

Note. N = 207. Values reflect the average across 7 years of data. * p < .05, two-tailed.

Figure 1 through Figure 3 reflected extrapolated data in regard to the input variables in FSA R%H, FSA M%H, and DSGs. Figures 4 and 5 show the output variables of the teacher instructional personnel evaluation ratings in the Highly Effective and Effective categories. The data were ascertained from a seven-year FLDOE analysis of two comparable K-12 public school districts.

Figure 1

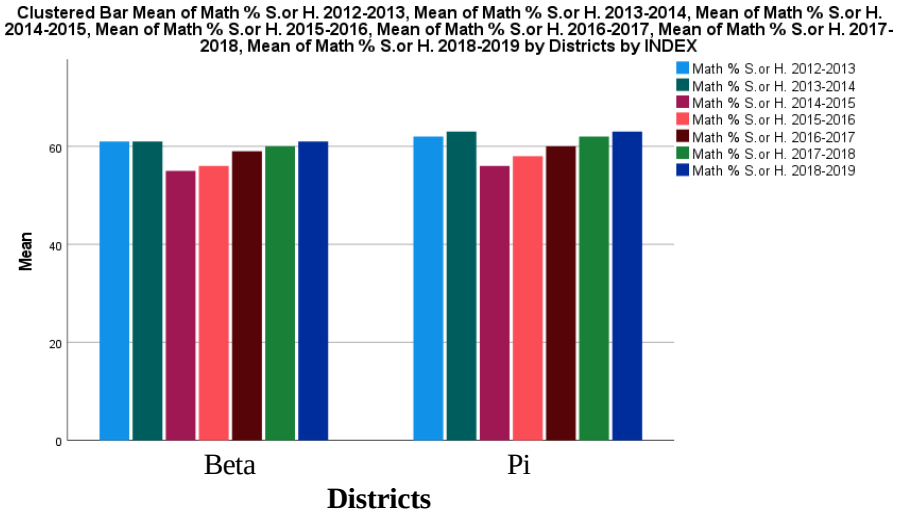
FSA English in the Domain of Reading % Satisfactory or Higher (x)-Input Variable Comparison Between Two K-12 Public School Districts (7-Year Analysis) - Clustered Bar Mean



Note. Table created in SPSS from selected district data retrieved electronically from the FLDOE. Copyright 2012-2019.

Figure 2

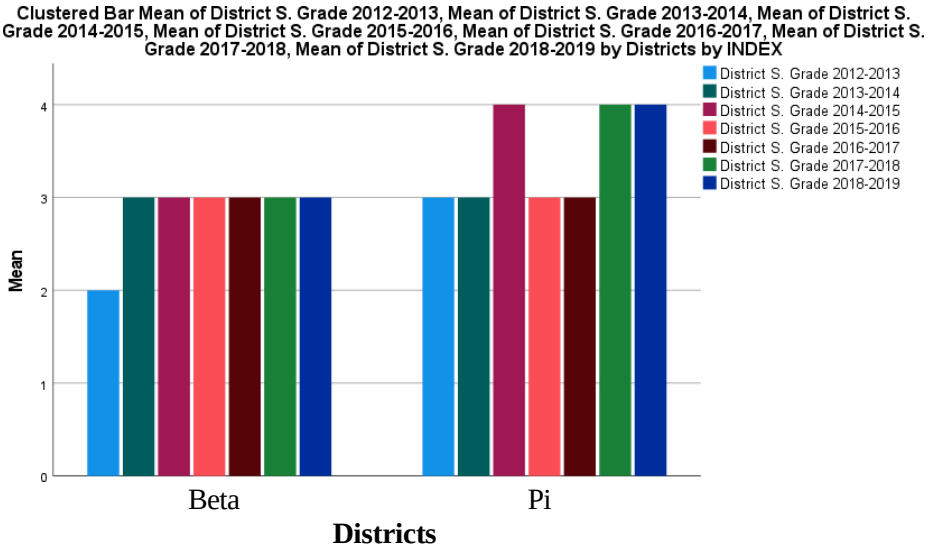
FSA Mathematics % Satisfactory or Higher (x)-Input Variable Comparison Between Two K-12 Public School Districts (7-Year Analysis) - Clustered Bar Mean



Note. Table created in SPSS from selected district data retrieved electronically from the FLDOE. Copyright 2012-2019.

Figure 3

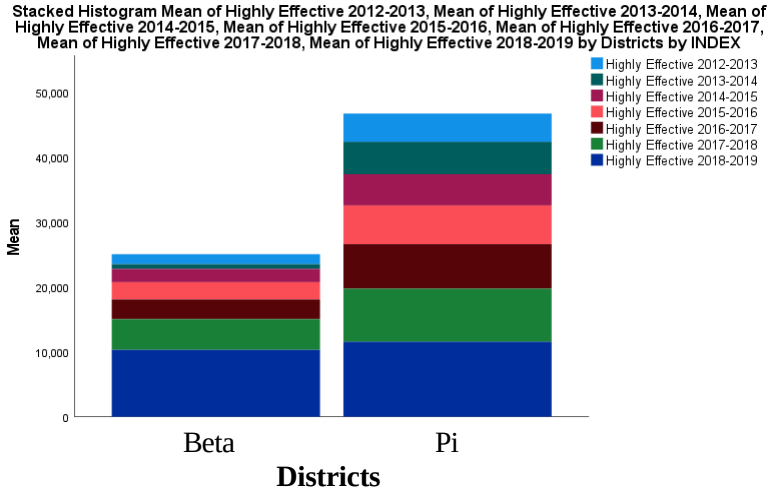
District School Grades (VAM/LGM Commensurate Assessment) for (x)-Input Variable Comparison Between Two K-12 Public School Districts (7-Year Analysis) - Clustered Bar Mean



Note. Table created in SPSS from selected district data retrieved electronically from the FLDOE. Copyright 2012-2019.

Figure 4

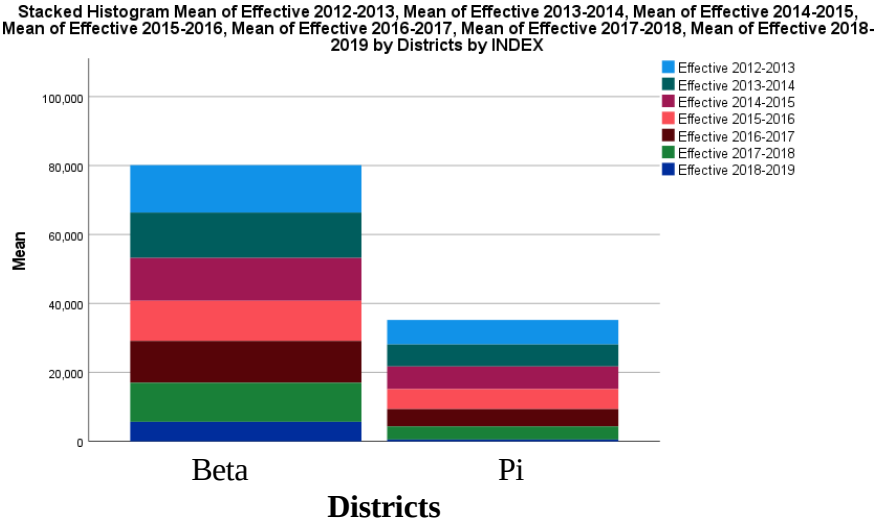
Highly Effective Teacher Instructional Personnel Evaluation Rating (y)-Output Variable Comparison Between Two K-12 Public School Districts (7-Year Analyses) - Stacked Histogram Mean



Note. Table created in SPSS from selected district data retrieved electronically from the FLDOE. Copyright 2012-2019.

Figure 5

Effective Teacher Instructional Personnel Evaluation Rating (y)-Output Variable Comparison Between Two K-12 Public School Districts - (7-Year Analysis) - Stacked Histogram Mean



Note. Table created in SPSS from selected district data retrieved electronically from the FLDOE. Copyright 2012-2019.

Chapter Summary

Since no statistically significant differences of the means between both districts can be determined to exist in the input variable categories of FSA (R%H and M%H) and DSGs between both large comparable urban based K-12 public school districts, then the question must be posed as to why there is such a drastic difference of the means of the two output variables between both K-12 public school districts in terms of the percentage of teaching instructional personnel rated THER versus TER when the FSA data (R%H and M%H) and DSGs are annually either the same percentage score or within two percentage points of each other? Both of these K-12 public school districts have annual billion dollar budgets. The School District of Beta County for the 2020-2021 academic school year has a budget of $4.5 Billion Dollars (BSD, 2021) and the School District of Pi County for the 2020-2021 academic school year has a budget of $3.8 Billion Dollars (Pi School District, 2021). Each K-12 public school district may place the weights differently in regard to teacher instructional personnel evaluation data (Beta – 50% Teacher Evaluation (IP), 35% Student Assessments and 15% VAM/LGMs (DSGs – a commensurate approximation); Pi – 33.3% Teacher Evaluation (IP), 33.3% Student Assessments and 33.3% VAM/LGMs (DSGs – a commensurate approximation) but since the input variables of FSA scores (R%H and M%H) and the VAM/LGMs (DSGs – a commensurate approximation) do not exhibit substantial statistically significant differences of the means between districts in comparison, then the remaining input variable of the teacher evaluation (IP) score has to be scrutinized. This teacher evaluation information is inputted by either the principal or the assistant principal at each school site. Therefore, qualitative information (what the reviewer observes in the classroom) is

being given a numeric code by the observer and it is transformed into a quantitative entity. Why is there such an imbalance if the three input variable domains are relatively equivalent? Based upon the analyses of the tested input variables that impact merit pay within the FSRP and FDAP in the two K-12 public school districts, the output variables (THER and TER) of the teacher evaluation are also impacted by the remaining input data created through the use of the iObservation ® tool (the IP score). It does not reflect the tested data evidence produced by the input variables of the FSA tests (R%H and M%H) and the VAM/LGM scores (DSGs). The numeric output from the iObservation ® tool is based upon subjective qualitative input and can therefore be partisan. Moreover, when an analysis of the Descriptive Statistics of the tested input variables for the K-12 public school districts was conducted, it may be surmised that the Marzano Evaluation Tool iObservation ® is being weaponized by either modifying, adjusting, or transmuting the teacher IP evaluation score thereby categorically impacting teacher placement on merit based pay-for-performance salary schedules.

CHAPTER 5. DISCUSSION AND CONCLUSIONS

A critically comprehensive historical understanding of the relevance and contextual role that Educational Leadership has played in constructing, framing, and maintaining our culture and society is essential to producing Transformative Educational Leaders in the 21st century. Murnane and Cohen (1986) argued that merit pay and education do not mix because teaching is difficult to evaluate. Without clear measures and criteria for judging success, decisions about rewarding performance are, at best, subjective and, at worst, unworkable. They also suggest that merit pay leads to the potential for dysfunctional behavior: "teachers may end up focusing on particular tasks or students that are rewarded by a merit pay plan at the expense of other important tasks or goals" (p. 4). By contrast, Ballou (2001) has argued that "there is nothing inherent in teaching and schooling that makes merit pay a poor fit" (p. 53). However, neither of the aforementioned arguments function in a singular linear capacity in the field of Education. Dixit (2002) asserts that "the complex and multidimensional tasks of teaching, the multiple goals that are vague and poorly observed, coupled with the presence of multiple stakeholders create minimal incentives in the public school sector" (p. 696).

Restatement of Purpose

The purpose of this study was to analyze teacher pay across two similar K-12 public school districts within the State of Florida. The selected districts served as a substantial indicator of the teacher evaluation methodology utilized in K-12 public education throughout the entire state. This research study sought to determine if the

districts maintained comparable input variables of the student assessment scores in English in the domain of Reading % Satisfactory or Higher (R%H), Mathematics % Satisfactory or Higher (M%H), and District School Grades (DSGs) which represent commensurate VAM and LGM scores. Furthermore, the study also endeavored to determine if the districts produced comparable output variables of the teacher evaluation ratings in the domains of Highly Effective (THER) and Effective (TER) if the input variables were comparable as well.

Quantitatively this means that this study analyzed if there was a statistically significant difference of the means in the (x)-input variable domains of FSA data in R%H, M%H, and DSGs which are components of the Florida Accountability Programs: the FSRP and the FDAP that fund and impact merit pay salary schedules. Additionally, this study endeavored to investigate if any percentage differences of the means existed in the (y)-output variables of teacher instructional personnel evaluation ratings in the THER and TER category domains. Over the past seven academic school years (2012-2019), this study endeavored to understand the output variable (teacher instructional personnel evaluation ratings – [THER] and [TER]) results which determine K-12 public school teachers pay differentials between these two K-12 public school districts (i.e. Beta and Pi) which constituted the foundation of the catalyst that necessitated this research.

Review of Methodology

This quantitative study included statistical tests and analyses that were utilized to respond to five research questions and to test five null hypotheses. In order to ascertain a more holistic perspective of this issue, the input and output variables were identified for data collection, testing and analyses. Archival statistics for the five input and output

variables over a 7-year period (2012-2019) were retrieved electronically from the FLDOE databases via the department's site.

During the seven-year longitudinal study, the input variables of FSA Data (R%H and M%H) and DSGs which are components of FSRP and FDAP served as the input variables. The output variables of Teacher Instructional Personnel Evaluation Ratings Data (THER and TER) served as the dependent-overlapping variables to determine the difference in correlations. Independent Sample t-tests were executed on the input variables of FSA (R%H and M%H) and DSGs (a commensurate approximation to VAM/LGM scores). The output variables of instructional teacher evaluation ratings in the categories of THER and TER were also analyzed. As an addendum, Bivariate Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were conducted to ascertain information regarding relationships and the difference in correlations that existed in the input and the dependent-overlapping output variables within each district and between both districts. Independent Samples t-tests and empirical data analyses were done upon the output variables of Teacher Personnel Evaluation Ranking Data for the domain designations of THER and TER. In this study, these variables were tested and analyzed because all are directly connected to K-12 public school Merit Pay salary schedule placement determinations. The data were stored in analytical computer software programs for statistical and empirical analyses.

The Independent Samples t-tests were performed to compare the means and standard deviations of average percentage scores in the input variables of (R%H, M%H, and DSGs) and the output variables (THER and TER). The schools within each K-12

public school district, Beta and Pi, served as the unit of analysis for these tests. The districts received an annual letter grade for each of the academic school years analyzed. Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts were also executed to obtain supplementary information regarding the relationships of the input and output variables. The output variables were tested and analyzed based upon the annually reported data on the Instructional Teaching Staff by every K-12 public school within each district. In this study, the most current information was analyzed as the FLDOE Emergency Order No. 2020-EO-1 concluded that the Spring K-12 statewide assessment test administrations for the 2019-20 school year were canceled due to the pandemic outbreak of the Coronavirus and accountability measures reliant on such data were not calculated for the 2019-2020 school year (FLDOE, 2020c).

Summary of the Findings

The following section of the paper presents the results of the analyses and the findings of the study (the input variables that contribute to merit pay through the Florida Accountability Programs [FSRP and FDAP] and the output variables that determine placement upon the pay-for-performance salary schedules) in relation to the existing literature. Morreale et al. (2017) assert that it is critical that scholars are able to communicate the nuances of what occurs within the world in which we live. If oversight and transparency are not embedded within an organizational framework, it will impact content and pedagogy thereby impeding those who comprise it their future personal and professional success.

Finding 1 – Affect of the Input Variables

The first three statistical analyses accepted the null hypotheses ($H_0$1, $H_0$2. and $H_0$3). The tests did not reveal any statistically significant differences of the means of the input variables of the FSA (R%H and M%H) and DSGs. Yet, the last two statistical analyses of the null hypotheses ($H_0$4 and $H_0$5) did show significant differences in the means of the teacher rating categories of Highly Effective and Effective.

The DSGs of the Beta District were negatively correlated to each of the FSA (R%H and M%H) within and between both districts. None of the aforementioned correlations were statistically significant. The Pi District DSGs were also shown to be negatively correlated with FSA (R%H and M%H) in district and between districts. None of the correlations were statistically significant. Moreover, for the Independent Samples t-tests that were run on the input variables (English, Math, and DSGs) and the output variables (Highly Effective and Effective), the Levene's Test for Equality of Variances showed significance levels above .01, therefore, the variances were assumed to be approximately equal for the input and output variable data analyzed.

Finding 2 – Influence of the Output Variables

The output variables of the teacher instructional personnel evaluation ratings of Highly Effective and Effective were shown through quantitative tests and empirical data analyses to have statistically significant differences in the percentage of teachers rated in the THER and TER categories between both districts. Additionally, Intercorrelation Matrices, Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts of the input to the output variables showed that there were differences in the correlations that were statistically significant in each of the FSRP and the FDAP input

variables of FSA (R%H and M%H) and DSGs in relation to the output variables of the teacher evaluation ratings for THER and TER in each district and between both K-12 public school districts.

Finding 3 – The Relationship Between the Input and Output Variables

The statistical tests served to accept the three null hypotheses ($H_0$1, $H_0$2 and $H_0$3) associated with the input variables. The statistical tests, however, showed significant differences in the final two null hypotheses ($H_0$4 and $H_0$5) associated with the output variables of Highly Effective and Effective. The Intercorrelation Matrices, the Independent Correlation Contrasts, and Overlapping Dependent Correlation Contrasts revealed the difference in correlations and served as an addendum to provide information regarding the relationships between the input and output variables within and between both districts.

The Intercorrelation Matrices revealed correlations that were statistically significant between the input variables, English and Math, within each district and between both districts. However, the Beta and Pi DSGs revealed negative correlation with no statistical significance with all of the FSA variable categories (English and Math). An argument can be made through the tested FLDOE data that since DSGs are used as a commensurate approximation of VAM/LGM scores, this validates the argument that the VAM/LGM (DSG) scores do not reflect increases in student achievement. Furthermore, the Highly Effective Rating was negatively correlated to the Effective Rating Domain within the Beta and Pi Districts but the correlation was not significant at the .05 level (2-tailed).

The Independent Correlation Contrasts between the two K-12 Public School Districts on the difference in correlations between the (x)-Input Variables (English, Math and District School Grades) and the (y)-Output Variables (Highly Effective and Effective) over the 7-Year Study (2012-2019) revealed that the p-value for all of the correlations were significantly different. The nature of these correlations was that English and Math were more strongly correlated in the Pi District than in the Beta District. The Overlapping Dependent Correlation Contrasts revealed the difference in predictive validity in the Beta and Pi Districts. Many of the correlations were significant in relation to the teacher evaluation rating in both the K-12 public school districts. Figures 1-3 represented the tested input data (R%H, M%H, and DSGs) and Figures 4-5 represented the analyzed output data (THER and TER).

This quantitative study showed through the empirical data testing that each of the input variables (R%H, M%H, and DSGs) contained within the FSRP and FDAP directly pertain to the merit pay-for-performance salary schedules for K-12 public school teachers have been in consistent alignment during the seven-year longitudinal study. However, the output variable empirical data of the K-12 public school teacher instructional personnel evaluation ratings (THER and TER) were not in alignment with the input variable data.

Impact of the Relevant Input and Output Variables In Relation To Merit Pay

The primary variables that are utilized to structure the merit pay salary schedule for K-12 public school teachers in Florida are integral in creating a framework that encourages recruitment and maintenance of a highly qualified educator base. The rating category designations of Highly Effective and Effective, which are the only two designations that provide monetary compensation to an educator's base pay, serve to

show every educator in the system what is deemed valuable, considered worthy, and ultimately who is chosen to receive compensation for the scholastic efforts of every educator within the system and for those who are considering entrance into the profession.

Impact of the Input Variable - Student Assessments (English & Mathematics)

The Independent Samples t-tests that were run on the K-12 public school assessments in FSA R%H and M%H in both districts revealed that there were no statistically significant differences of the means between the selected input variables (English and Mathematics) of the FSRP and the FDAP for each K-12 public school district that would impact components of the output variables of teacher evaluation ratings (Highly Effective and Effective). The aforementioned designations serve to determine placement of K-12 public school teachers on merit pay-for-performance salary schedules. In terms of the empirical research questions and null hypotheses that were posed in this study regarding the student assessments:

1. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA English in the domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts?

$H_0$1. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA English in the domain of Reading % Satisfactory or Higher (R%H) between two large urban based Florida K-12 Public School Districts.

2. Is there a statistically significant difference of the means between the (*x*)-input variable of the FSA domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts?

$H_0$2. There is no statistically significant difference of the means between the (*x*)-input variable of the FSA Domain of Mathematics % Satisfactory or Higher (M%H) between two large urban based Florida K-12 Public School Districts.

The testing and analyses showed that there was no statistically significant difference of the means between each FSA student assessment variable (English and Mathematics) between the two large urban based Florida K-12 Public School Districts.

Impact of the Input Variable - DSGs In Relation To VAM/LGM Scores

The Independent Samples t-tests that were run on the K-12 public school DSGs in both districts revealed that there were no statistically significant differences of the means between the selected input variable, DSGs, reflected in the FSRP and FDAP for each K-12 public school district that would impact components of the output variables of teacher evaluation ratings (Highly Effective and Effective). This input variable serves to determine placement of K-12 public school teachers on merit pay-for-performance salary schedules. In terms of the empirical research questions and null hypotheses that were posed in this study regarding the district school grades:

3. Is there a statistically significant difference of the means between the (*x*)-input variable of the DSGs between two large urban based Florida K-12 Public School Districts?

$H_0$3. There is no statistically significant difference of the means between the (*x*)-input variable of the DSGs between two large urban based Florida K-12 Public School Districts.

District School Grades are commensurate with an approximation to the VAM and LGM scores. Since the FLDOE does not release individual K-12 public school teacher VAM and LGM scores, the DSGs were the most appropriate comparative score to utilize when comparing these two districts. These are also the principle variables utilized by the Florida K-12 Public School Accountability Programs (FSRP and the FDAP) that directly impact merit pay.

Impact of the Output Variable - Highly Effective Teacher Rating

An Independent Samples t-test was conducted to compare the output variable of the Highly Effective Teacher Rating data of all Instructional Teachers between two K-12 public school districts. The following empirical question and null hypotheses were presented:

4. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (*y*)-output variable of the Highly Effective (THER) Category Rating Designation between two large urban based Florida K-12 Public School Districts?

$H_0$4. There is no statistically significant difference in the percentage of the mean between the (*y*)-output variable of the Highly Effective (THER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

When the tests and analyses were run, it was determined that over the

7-year longitudinal study (2012-2019), the significance level (*p*) for the Independent Samples t-test revealed statistically significant differences in the means that addressed the fourth null hypothesis ($H_0$4). Consequently, there is a statistically significant difference of the means in the percentage between those rated in the Highly Effective Category Rating Designation between two large urban based Florida K-12 Public School Districts.

Impact of the Output Variable - Effective Teacher Rating

An Independent Samples t-test was conducted to compare the output variable of the Effective Teacher Rating data of all Instructional Teachers between two K-12 public school districts. Over the 7-year longitudinal study (2012-2019), the following empirical question and null hypotheses were posed:

5. Is there a statistically significant difference in the percentage of K-12 public school instructional teachers rated in the (*y*)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts?

$H_0$5. There is no statistically significant difference in the percentage of the means between the (*y*)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

Over the 7-year longitudinal study (2012-2019), it was determined that the significance level (*p*) for the Independent Samples t-test revealed statistically significant differences in the means that addressed the fifth null hypothesis ($H_0$5). Consequently, there is a statistically significant difference of the means in the percentage between the

(*y*)-output variable of the Effective (TER) Category Rating Designation between two large urban based Florida K-12 Public School Districts.

In both K-12 public school districts (Beta and Pi), the IP component of the teacher evaluation must be further investigated. Even though each district weighs each teacher's IP evaluation score component differently, if the input variables were correspondingly commensurate during this 7-year study, then the IP domain which is subjective must be assiduously examined. Increased transparency must occur to ensure that any bias or inequity does not continue to suffuse throughout the evaluation process thereby serving as an impediment for Highly Effective teachers to be recognized by the system. According to Marzano Research (2022), the "view of improvement assumes that the people involved have the ability and the will to learn what is necessary to improve student academic and social emotional outcomes" (p. 1). However, the iObservation ® tool was initially designed for an educator's self-reflection. The iObservation ® Site maintains that the Marzano Protocol (2019) on iObservation ® is part of the Marzano Suite for Connecting Teacher Growth to Student Achievement. It was not supposed to be linked to monetary incentivization as it could be manipulated if it was not utilized correctly. Dee and Wyckoff (2015) state that "the single salary schedules commonly used in U.S. public school districts compensate teachers according to tightly structured rules. They assert that such rigid and misaligned compensation systems cannot adequately attract and retain a high-quality teacher workforce" (p. 267).

Merit-Based Pay and Teacher Instructional Personnel Evaluations/FSA Student Assessments/VAM/LGM Score Relationship to DSGs

Shelly (2008) maintained that RTTT complex design reflected the Obama administration's need to navigate a difficult political situation, a deep "inter" and "intra" party division over school reform—as well as a difficult institutional situation—the limited capacity of federal and state education agencies to push reform down to the school level. "This is what might be called the 50/14,000/130,000 problem in American education reform—we have 50 different state education systems that collectively contain approximately 14,000 public school districts and almost 130,000 schools" (p. 447). States have developed vastly different education systems, and tremendous variation in school quality exists within and among states. "Although the United States now has clear national goals in education, it lacks a national *system* of education within which to pursue these goals, and the federal government can only indirectly attempt to drive reform through the grant-in-aid system" (Shelly, 2008, p. 454).

Cohen and Moffitt (2010) have observed, this has greatly limited the federal government's ability to effect change in education and stymied its pursuit of educational equality. Although the federal government can sometimes use incentives to coerce states into adopting certain policies, it has struggled to get states to implement them faithfully or effectively, and they have, therefore, often failed to achieve their aims (Manna, 2010).

Thus, a question must be posed that focuses upon how teacher leaders can move up in this system with the barriers in place that are imbued within this type of observational methodology which is utilized in teacher instructional personnel evaluations. According to McGuinn (2012), lawsuits have been filed against the State of

Florida claiming that the teacher evaluation system is unfair because it partly rates their job performance on test scores of students they do not know and subjects they do not teach. “Since 2009, thirty-six states and the District of Columbia have required that teachers be evaluated in part based on student scores on standardized tests. The idea received a boost because of Obama’s administration policies, particularly Race to the Top” (McGuinn, 2012, p. 136). According to Harrison and Cohen-Vogel (2012), Senate Bills 6 and, later, 736, which became law in 2011, put forward ideas for substantively altering the teaching profession in the State of Florida. Under the new law, “teachers, traditionally compensated through pay scales that reward advanced education or experience, faced a new system in which performance-based evaluations drive salary increases” (p. 519). These systems must base at least 50% of a teacher’s evaluation on his or her VAM or LGM value added to student achievement (Florida Senate, 2011). In addition, Senate Bill (SB) 736 eliminated the State’s previous policy, which required that districts award “tenure” to teachers after 3 to 5 years of service by offering them long-term professional service contracts (Florida Contracts with Instructional Staff and Supervisors Act [2010] – Florida Statute §1012.33 €). Under the new law, all K-12 public school teachers hired after July 1, 2011, were offered annual contracts that expire at the conclusion of each year, regardless of their length of service.

In the Public School District of Beta County, the Educational Professionals (EP) Unit of the Beta Teachers Union created a MOU which is not permanently binding and must be fully negotiated into contract language for it to continue. It contains the following IP Adjustments beginning in 2018-2019:

- IP will be worth 50% of the overall evaluation.

- Move Applying/Accomplished from 3.0 to 3.25.
- The parties agree to the following Deliberate/Professional Practice (DP) Adjustments beginning in 2018-2019 – Deliberate/Professional Practice (DP) will be worth 15% of the overall evaluation.
- The parties agree to the following Student Performance Adjustments beginning in 2018-2019: For 4th-10th grade ELA, 4th-8th grade Math, and 8th-9th grade Algebra I teachers, the District will continue to use a Local Growth Model (LGM) instead of the State's VAM. Currently, this MOU has not been made permanent. (Beta Teachers Union, 2021)

In the Public School District of Pi County, salaries are based on training and experience. In accordance with the Classroom Teachers Association (CTA) contract, teachers may be given credit for up to 25 years of prior full-time creditable school teaching experience, to be granted on an equitable basis according to the New Hire Salary Placement Schedule. Experience must be verifiable as a certified teacher (Pi Classroom Teachers Association, 2021). The Performance Salary Schedule is established pursuant to the Florida Personnel Act (2021) which can be found in the Florida Statute §1012.22(1)(c)(1)(d) and includes, but is not limited to the following Student Success Act legislative provisions:

- The annual salary adjustment under the Performance Salary Schedule for an employee that is rated as 'Highly Effective' must be greater than the highest annual salary adjustment available to an employee of the same classification through any other salary schedule adopted by the District (Florida Personnel Act [2021] - Florida Statute §1012.22(1)(c)(1)(d)).

- The Performance Salary Schedule shall not provide an annual performance salary adjustment for an employee who receives a rating other than 'Highly Effective' or 'Effective' for the evaluation year (Florida Personnel Act [2021] - Florida Statute §1012.22(1)(c)(1)(d)).
- Teachers new to the District or Teachers that are rehired after a break in service must successfully complete a one (1) year probationary annual contract before becoming eligible for a one (1) year non-probationary annual contract. A Teacher may be terminated at any time during the probationary period as a regular probationary teacher. (Florida Personnel Act [2021] - Florida Statute §1012.22(1)(c)(1)(d)).

According to the Pi District CTA (2021) and the Pi District Official Site (2021) that is responsible for teacher instructional personnel salary contract negotiations:

> Effective July 1, 2020, the Grandfathered "Open Range" (minimum-maximum) Salary Schedule has the beginning base salary of $47,500 and a top annual base salary of $90,029. The Parties agree that the only differences between the "Open Range" Grandfathered Salary Schedule and the "Open Range" Performance Pay Salary Schedule is that employees having a Continuing Contract (CC) or a Professional Services Contract (PSC) with the District need not give up or forfeit their CC or PSC status in order to receive Performance Pay Increases that are negotiated by the Parties, and that on the date the new negotiated salary adjustments become effective, an employee rated "Highly Effective on the "Open Range" Grandfathered Salary Schedule will be paid one dollar ($1) less than an employee rated "Highly Effective" who is on the Performance Pay Salary Schedule, and an employee rated "Effective" on the "Open Range" Grandfathered

> Salary Schedule will be paid one dollar ($1) less than an employee rated "Effective" who is on the Performance Pay Salary Schedule. Due to the COVID-19 Pandemic, for SY 20-21, no performance ratings are available to use; therefore, any salary increases will be the same on both the Grandfather "Open Range" and Performance Pay Salary Schedules. (Pi District Schools Compensation Department, 2021; Pi CTA Collective Bargaining Agreement Contract, 2021, Article VIII – Compensation and Benefits, p. 116)

Implications of Linking Merit Pay to Teacher Performance

According to the NCES, the most recent data revealed that in the academic school year 2019-2020, the U.S. K-12 public school student enrollment consisted of approximately 51.5 million students (National Center for Education Statistics, 2021). "The U.S. K-12 private school student enrollment embodied approximately 4.7 million students" (Broughman et. al., 2021, p. 2). This data confirmed that approximately 92% of all students attend a K-12 public school while 8% attend a K-12 private school in the United States. The existential crisis that is occurring in the State of Florida and throughout the nation concerns how Highly Effective teachers will be recruited and maintained within the educational infrastructure that is currently in place. Horace Mann was known as the "Father of the Common School". His ideological paradigm serves as the foundation for the modern-day public school. In 1837, Mann presented a report that stated:

> A nation cannot long remain ignorant and free. No political structure, however artfully devised, can inherently guarantee the rights and liberties of citizens, for freedom can be secure only as knowledge is widely distributed among the

> populace. Universal education is the only foundation on which a government can securely rest. (Mann & Cremin, 1979, p. 7)

Florida Assessment of Student Thinking (FAST) High-Stakes Progress Monitoring Test and Florida Benchmarks for Excellent Student Thinking (BEST) Standards

It must be reiterated that on September 14, 2021, Florida Governor Ron DeSantis, announced a legislative proposal that will eliminate the common-core based, end-of-year, high-stakes FSA in the Spring of 2022 and create the new Florida Assessment of Student Thinking (FAST) plan that will go into effect in 2023-2024 with a gap year in 2022-2023 to obtain baseline data. This FAST Progress Monitoring Test will monitor student progress and foster individual growth through three tests that will occur throughout the school year. By creating the FAST Progress Monitoring Test, Florida will become the first State in the nation to fully implement progress monitoring instead of end-of-year standardized testing and fully eliminate common core (FLDOE, 2021f). With the implementation of this new testing paradigm, it does not actually end high-stakes testing accountability and testing in the Spring term will still occur.

The FSA will morph into FAST and instead of a singular high-stakes test at the end of the Spring semester there will be *three* high-stakes tests throughout the academic school year still culminating with a final high-stakes test in the Spring semester for all K-12 public school students that are tied to school accountability, teacher pay, and graduation. It is important to acknowledge the current trajectory of how the State of Florida intends to implement the upcoming FAST Progress Monitoring Tests. Common Core will be changed to the BEST Standards. The three high stakes tests will be created by the Cambium Learning Group, Inc. (FLDOE, 2021d) which is owned by Veritas

Capital (Business Wire, 2018), a New York-based multi-billion dollar private equity firm. Furthermore, the FLDOE Commissioner of Education, Richard Corcoran, approved Lexia Learning for English Language Arts (ELA) Intervention Courses at the Elementary Level and Power Up Literacy Intervention Courses for Grades 6 and Up (Cambium Learning Group, 2021). Lexia Learning is branded software produced by the Cambium Learning Group Company (Cambium Learning Group, 2021) which is also owned by Veritas Capital (Business Wire, 2018). Veritas Capital (2020) owns all of the subsidiary companies that is under the auspices of the Cambium Learning Group.

Private equity firms, like Veritas Capital, are not required by law to release a multitude of "private" corporate information. Lossen (2007) maintains that "there is a lack of systematic information in regard to portfolios and strategies utilized by Private Equity (PE) Firms" (p. 3). This type of ambiguity severely hinders and impedes the transparency required by K-12 schools, teachers, and parents who would need to know essential information regarding the specifics of the FAST Progress Monitoring Test and its vetting and reliability protocols since these FAST Progress Monitoring high-stakes tests will be "linked to school accountability, student graduation as well as merit pay" (FLDOE, 2021d). However, the FLDOE (2021f) opened Rule 6A-1.09401, Florida Administrative Code, for rule development to adopt or revise student academic standards for several content areas. As such, the Cambium Learning Group has been designated to create the high-stakes FAST Progress Monitoring Tests and have input on the BEST Standards in academic content areas for the State of Florida (FLDOE, 2021f). Additionally, Alchemer, formerly known as SurveyGizmo, was used by FLDOE to obtain feedback from K-12 public school personnel (i.e. Teachers and Administrators), parents,

and community members (FLDOE, 2021f). Alchemer is not a publicly traded but rather a privately held company (Alchemer, 2021).

Information from the Florida Department of Financial Services [FLDFS] (2021) regarding the Florida Accountability Contract Tracking System delineates that currently close to $160 million dollars is being allocated yearly by the FLDOE to the Cambium Assessment, Inc. in regard to current FSA Tests which are administered annually before the next iteration of student assessment, the FAST Test, is implemented (FLDFS, 2021).

Impact of Merit Pay System Schemes on Educational Leadership Growth

Descartes and Buchenau (2016) held that "it is morally impossible that there should be enough different devices in a machine to make it behave in *all* the occurrences of life as our *reason* makes us behave" (p. 37). According to Wertz (1998), in modern times, this is the way Descartes would have differentiated humans from varied forms of artificial intelligence. If the assumption is that humans use reason, it must be noted that authors of various disciplines such as philosophy, policy, sociology, psychology, anthropology, and education have discussed the meaning and importance of maintaining a socially just society (Brooks et al., 2015; Fraser, 2000; Jean-Marie et al., 2009; Oplatka et al., 2014; Preckler et al., 2012) and the ways to reach such a society is through education (Berkovich, 2014).

Thomas and Lang (1937) state that “Education is all that stands between our present standard of living and that of the most primitive savage” (p. 60). In terms of Educational Leadership, the connection to Descartes’ most well-known statement must be acknowledged. In his *Discourse On Method* (1641), his statement, “I think therefore I am” became a fundamental element in Western philosophy. Bode (1921) supported this

belief when he stated, “Life is more than vocation, culture, knowledge, and citizenship. All these interests are interwoven in an endless tangle, so as to give some color to the notion that the educational problem lies in the concept of growth” (p. 12).

Before he became the second president of the United States, John Adams stated, “All the children of the inhabitants, the rich as well as the poor, have a right to go to these public schools” (Hinsdale, 1898, p. 46). In his capacity as the second American President, he was able to orchestrate political change and support systemic policies that would serve to educate children from both rich and poor backgrounds (Skowronek, 2003).

In no area of school operations has the mounting contentiousness been more evident than finance. When coffers are full, school systems may be able to placate most groups making educational demands, but when resources dry up and cuts must be made, educators quickly get a lesson in who has political influence and who does not possess it (Duke, 2005). The first half of the 20th century was one of stability as educational leadership programs were dominated by practitioners. “That period was succeeded by a quarter of a century during which leadership preparation programs were driven by specialists who presented social science frameworks” (Milstein, 1999, p. 538). No longer can school system leaders count on a reasonable level of agreement when changes are proposed. Most reform initiatives bring out well-organized groups of advocates and opponents (Duke, 2005). According to the Webers’ research (1955), guiding principles of leadership should be conceived as the formulation of statements of habits of behavior which have been found to engender the growth of people in the direction of increased biological vigor, mutuality, and use of intelligence in solving problems and creativity.

Impediments Within Merit Pay Systems

In relation to the context of this research, which focused upon the K-12 public school pay-for-performance schemes in education, according to Castilla (2012), "racial minorities consistently receive lower performance ratings than white employees after controlling for job, work unit, and supervisor fixed effects" (p. 530). Recent empirical studies have suggested that "workplace inequality remains even following the adoption of employer processes meant to increase diversity and reduce ascriptive inequality, such as diversity policies" (Kalev et al., 2006, p. 592). In these merit-based pay-for performance plans, "the evaluations determine the training and development opportunities, transfers, demotions, and terminations of employees" (Greenhaus et al., 1990, p. 66). Moreover, those who study the current work arrangements have raised equity concerns about the implementation of such practices and routines (Castilla, 2008; Dencker, 2008; Fernandez-Mateo, 2009; Manning & Swaffield, 2008).

According to Verger et al. (2014):

> The idea is that employees will comply voluntarily with norms and rules, the more they are convinced of the legitimacy of the rule. The legitimacy of a rule can result from beliefs in the moral validity of the norm itself, but it can also result from beliefs in the validity of the procedure by which the rule had been worked out from the micro-mechanism underlying this type of social steering that should concern those involved in this process. (p. 382)

They contend that the formalization of such practices masks inequality in the distribution of rewards (Elvira & Graham, 2002; Reskin, 2000).

The Department of Education's Inspector General (1996) repeatedly perseverated that "financial programs within education are the most vulnerable to fraud, waste, and abuse" (p. 8) in part due to the multitude of entities that assist in administering these programs. Oversight from external sources should be explored. According to Schwartz and Fayer (2006), "the equality required by democracy can easily be experienced as thin in contrast with the robust forms of substantive equality associated with that of distributive justice" (p. 294). All stakeholders must have a voice in the process.

Merit Pay and the Teaching Profession

According to Kelly and Hill (2002), "despite the Equal Pay Act of 1973, and the highest number ever of women working – over 70% – women still only earn 82% of the male wage; and these figures are based on full-time work" (p. 211). According to the National Center for Education Statistics (NCES) (2021):

> About 76% of PK-12 public school teachers were female and 24% were male in 2017–18, with a lower percentage of male teachers at the elementary school level (11%) than at the secondary school level (36%). Overall, the percentage of public school teachers who were male was 2% points lower in 2017–18 than in 1999–2000. (p. 1)

However, Laker and Davis (2011) assert that it is not men who ultimately benefit from the oppression of women, but capital and our economic system. Based upon the work from Gowlett and Rasmussen (2014), they state that "with such a conception of power and radical action, it is difficult to generate the idea of a policy process that involves collective action, and institution building" (p. 331).

Implications For Practice and Policy

If we look at gender inequity from a global perspective, additional insights could be ascertained. According to Cooray and Potrafke (2011), "if the education of girls is conducive to economic development, the self-preservation of political elites in these societies is a suggested explanation for gender and financial bias in education in government schools" (p. 271). Critics point at the ambiguous relationship between educational inputs and outputs and maintain that outcome-based accountability exemplified by high stakes testing does not narrow socioeconomic and racial gaps in scholastic achievement and educator productivity (Harris & Herrington 2006; Hursh, 2005; Mintrop, 2004). Browder and Cooper-Duffy (2003) also "question the appropriateness and effectiveness of outcome-based accountability with regard to students with disabilities" (p. 158), the impact on teacher creativity, and youth whose native language is not English (Alamillo et al., 2005).

Ballou and Podgursky's (1993) research show that the NEA believes that teacher instructional pay-for-performance salary schedules, such as merit pay, are inappropriate because of the complexity of the teaching-learning process (p. 51). Utilizing student achievement to assess teacher instructional performance is often resisted on the grounds that achievement is influenced by many factors beyond the instructors' control (Johnson, 1986; Hatry & Greiner, 1984). Merit pay plans limit awards to a pre-specified fraction of the teaching instructional personnel, a quota system that many deem unfair (Bacharach et al., 1985). This chapter discussed the findings of the study and recommendations. These recommendations fall in the areas of Policy and Practice, Recommendations for

Transformative Leadership and Policy Development, Recommendations for Future Research, and Researcher Reflections on the Findings from the Study.

Recommendations for Policy and Practice

According to Goldhaber et al. (2008), the fundamental assumption in a merit pay plan is that the district is interested in only one teacher activity that is labeled as *teaching* and in one outcome that is called *student achievement*. In reality, teachers pursue multiple goals that districts care about (i.e., increasing students' academic achievement, fostering their emotional and physical growth, preparing them for citizenship, etc.). Nevertheless, given today's policy and political interest in improving achievement for all students, "it is not unreasonable to portray a public school district as being principally interested in teaching activity that leads to student achievement outcomes" (Goldhaber et al., 2008, p. 268). Additionally, Hanushek (1979) asserts that from the teachers' perspective, all factors outside of their control are random. This abstraction reflects the reality that student outcomes depend on a host of factors that go beyond the teacher's "teaching" efforts.

According to Dewey (1917):

> A society not only continues to exist by transmission, by communication, but it may fairly be said to exist in transmission, in communication. There is more than a verbal tie between the words common, community, and communication. In order to form a community or society we must have aims, beliefs, aspirations, knowledge—a common understanding. Such things cannot be passed physically from one to another, like bricks; they cannot be shared as people would share a pie by dividing it into physical pieces. (p. 6)

Furthermore, Shields (2004) contends:

> Educational Leaders are expected to develop learning communities, build the professional capacity of teachers, take advice from parents, engage in collaborative and consultative decision making, resolve conflicts, engage in effective instructional leadership, and attend respectfully, immediately, and appropriately to the needs and requests of families with diverse cultural, ethnic, and socioeconomic backgrounds. (p. 109)

Hence, "authentic dialogue lies at the core of organizational learning, for without dialogue, individuals and groups cannot effectively exchange ideas, nor can they develop shared understanding" (Mazutis & Slawinski, 2008, p. 437). The Art and practice of teaching is not fully recognized in merit pay schemes. We must remain cognizant of the threats to academic and intellectual freedom. Bury (1913) wrote, "the struggle of reason against authority has ended in what appears now to be a decisive and permanent victory for liberty. In the most civilized and progressive countries, freedom of discussion is recognized as a fundamental" (p. 248). Our history is a living legacy that is dynamic and continues to evolve. As Butler (2015) notes, the construction of children is "neither a subject nor its act, but a process of reiteration by which both "subjects" and "acts" come to appear at all" (p. 9). However, in the 21st century, as the nature of teaching evolves under standards and accountability, and as market-based reforms introduce new personnel policies in education, "merit pay reforms may become increasingly attractive to districts and states" (Goldhaber et al., 2008, p. 285).

The quandary of ascertaining and maintaining high quality educators in a merit pay based system that persists is echoed by the insights of Dewey (1917) when he asserted:

> The necessity of teaching and learning for the continued existence of a society that we may seem to be dwelling is a truism. Life demands teaching and learning for its own permanence, but the very process of living together educates. It enlarges and enlightens experience; it stimulates and enriches imagination; it creates responsibility for accuracy and vividness of statement and thought. (p. 36)

Dewey is one of the architects of what it means to obtain an education in a democratic society (Maher, 1999). Dewey and current educational leaders do not place restrictions on the parameters of what it means to be an educator. Therefore, teachers must learn to determine students' learning needs, identify and implement activities and methods that will meet these needs, and reflect on the effectiveness of their implementations (Baker & Rozendal, 2019). Scholars have also argued that the merit pay plan schemes have allowed public sector employers to simultaneously shed responsibility and disempower the workforce (Wills, 2008).

Recommendations to Transformative Leadership and Policy Development

Giroux (1992) asserts that difficulties in educational leadership is associated with crises within a democratic government. Van Alfen (1993) maintains that others are concerned about the lack of leadership offered by school boards themselves or about the propensity of educators to adopt a series of reforms in rapid succession (Fullan, 2003), failing to empower either teachers or administrators. Dewey's (1938) comparative analyses of traditional practices found that schools contrasted with his critical and reflective methods on reconstructing knowledge for improving society. These constructs were, at heart, pathways for educators working towards social justice (Bogotch, 2002). Dewey's ideas were not reducible to convergent thinking, either-or polarized thinking,

standardization, or best practices. Yet today, in too many contexts throughout the world, perhaps most notably inside the school reform movements in the United States (i.e., NCLB/RTTT), educational leaders find themselves on an unsettling middle passage journey to what Dewey would call amoral actions and miseducative practices (Bogotch, 2012; Bogotch & Roy, 1997).

According to Hargreaves and Fink (2004):

> Sustainable leadership matters, spreads, and lasts. It is a shared responsibility, that does not unduly deplete human or financial resources, and that cares for and avoids exerting negative damage on the surrounding educational and community environment. Sustainable leadership has activist engagement with the forces that affect it, and builds an educational environment of organizational diversity that promotes cross-fertilization of good ideas and successful practices in communities of shared learning and development. (p. 10)

Thus, it is crucial to have an understanding and overview of those who have had a significant impact upon the field of Educational Leadership in order to make enlightened progress, change, and transformation in this field in the 21st century. In Plato's *Republic*, he speaks to the notion that justice is the greatest good of the soul, injustice its greatest evil (Kraut, 1997). Plato wrote, "What sort of knowledge is there which would draw the soul from becoming to being? All arts, sciences, and intelligences use in common, that which everyone first has to learn among the elements of education" (Republic – Book VI). The guardians of his society were to be true philosophers, passionately committed to reason, and trained in its rigorous educational application (Bluestone, 1987). Gutman

(1999) also forged connections with Plato's construction of an ideal society in that the importance of deliberation is also stressed. Plato (1955, Original work written 360 BCE) asserted:

> Deliberation is not a single skill or virtue. It calls upon skills of literacy, numeracy, and critical thinking, as well as contextual knowledge, understanding, and an appreciation of other people's perspectives. By cultivating these and other deliberative skills and virtues, a democratic society helps secure both the basic opportunity of individuals and its collective capacity to pursue justice. (p. 18)

According to Davidson (2012), Jean-Jaques Rousseau demanded absolute free play for the feelings and emotions. This was an essential component for those who wished to learn in a meaningful way. Rousseau believed that a sense of justice proceeds from an experience of the sense of injustice (Ferguson, 1984). Based upon Orwin and Tarcov's (1997) translation, Rousseau declared:

> The most brilliant thoughts can come into children's brains, or rather the best lines into their mouths, as diamonds of the greatest value might come into their hands, without either the thoughts or the diamonds thereby belonging to them. There is no true property of any kind at that age. (p. 32)

As someone who was immersed in Education, Rousseau thought it was important to encourage the emergence of the individual through creative means. According to Compayré (1971), Rousseau's life-long ambition was to show that through education, one can expect to guide new souls along the paths of a regenerated existence. Goldhaber et al. (2011) maintain that "these educational merit pay plans make teaching unattractive for high achieving people with technical skills and make difficult teaching assignments

unattractive for incumbents. This type of system does little to motivate and reward effective teaching" (p. 441). Reformers suggest that "teachers who do well, as well as those in unpleasant circumstances, those taking on special challenges, and those with critical skills ought to be pulling in very different base salaries" (Hess, 2014, p. 119).

Young (2014) argued that "industrialism compartmentalized and alienated students and school teaching faculty. It served to rob persons of their humanity, creativity, and intelligence because it made no room for it" (p. 321). A challenge for 21st century Educational Leaders is quality communication. This necessitates the merging of the fields of communication and educational leadership (Webb et al., 2012). Blount (2006) asserts that "those vested with power also were those most removed from the situations over which their judgment was directed, thereby minimizing the possibility of responsible decision-making" (p. 1085). Ultimately, this system thrives on isolation or division between people as well as between thinking and doing.

In relation to this research study regarding merit pay schemes in K-12 public education, Pearce et al. (1987) assert that "performance pay will not work where it is simply test-based, both because it is conceptually flawed and because it simply cannot be implemented in any meaningful way" (p. 169). Furthermore, "some districts will not have the patience and courage to review their practices and align purposes, goals, structures, practices, and methods of compensation" (Gratz, 2010, p. 17). In terms of social justice within the field of Education, it would require "actively working to change social institutions, political, and economic systems, and governmental structures that perpetuate unfair practices, structures and policies in terms of accessibility, resource distribution and human rights" (Toporek et al., 2006, p. 10).

May and Delston (2017) report that Article 26 of the United Nations Universal Declaration of Human Rights in regard to Education states, “Everyone has the right to education. It shall be directed to the full development of the human personality and to the strengthening of respect for human rights and fundamental freedoms” (p. 56). According to Cole (2018), it is necessary to make a distinction between equality and equal opportunities. Equal opportunities policies, in schools, seek to enhance social mobility within structures which are essentially unequal. They seek a meritocracy where people rise or fall on merit, but to grossly unequal levels or strata in society – unequal in terms of income, wealth, lifestyle, life chances, and power.

In terms of transformative leadership in Education, Munby and Fullan (2016) affirm:

> Many have worked for years in systems which are caught in a struggle between state and country level policy on the one hand, and the action or inaction of individual school districts on the other. Policy pushes in one direction, the profession pulls in another. The result is a type of friction which produces heat but not light: plenty of activity but not enough systematic change or improvement in outcomes. (p. 3)

According to Munby (2020), it is possible to mandate the changes and to reinforce implementation through “rigorous monitoring and high-stakes accountability, but this is expensive, is hard to sustain and is likely to have a negative impact on the attraction and retention of good teachers and leaders” (p. 147). Furthermore, the Webers (1955) believed that leadership is more than management. To refer to educational administration as leadership is “to assign it a more significant role than mere management. It is far too

limited a view to conceive of administration in strictly operational terms. Management is an important part of administration but it is not synonymous with leadership" (p. 57). Pisapia (2009) also maintained that a leader "looks for the windows of opportunity to open and they *run for daylight*" (p. 99). The leader would have to be cognizant of the elements and people that could deter their efforts. Some people would avoid this type of commitment to change "because they dislike the conflict that naturally comes with any ambitious team activity" (Kegan & Lahey, 2001, p. 86). It is not easy to articulate intent but a genuine strategic leader must look for the signs and find ways in which to anticipate what may occur at their educational organizations due to any changes that might take place.

Sanford et al. (2019) assert that educational leadership should be conceptualized and developed differently. Their research states:

> The world has been rapidly changing with new technologies, access to information, family structures, diversity and mobility, environmental sustainability and awareness of ecological challenges, to name a few. As a result, learners and their needs have changed. However, leadership roles often remain static, hierarchical, and linear, stuck in an industrialized model of management. (p. 4)

Leadership scholar Fairholm (2004) states that "we are caught in a Newtonian metaphor of mechanistic predictability and "the world of organizations has come to recognize the limitations of traditional management theories to describe fully the "hows" and "whats" of operating in a collective environment" (p. 370). Lunenburg and Ornstein (2022) specifically identified communication as the requisite skill of today's school leader.

"Tasks cannot be accomplished, objectives cannot be met, and decisions cannot be implemented without adequate communication" (p. 210). In the 21st century global village, "strong communication skills are vital for educational leaders around the world" (Webb et al., 2012, p. 88).

Moreover, according to Grogan et al. (2011), no federal or national organization, including the NCES, collects or reports annual administrative data by gender – let alone by gender and ethnicity combined – there is no easy way to compare the representation of women in educational leadership positions from year to year. Currently, the field relies on membership counts in professional educational leadership organizations, or intermittent surveys by the NCES to report the percentage of leadership positions in public and private schools held by women. Tyack and Hansot (1994) reported that "the absence of data had historical precedence in that data by gender became strangely inaccessible" (p. 31). A conspiracy of silence could hardly have been unintentional. Blount (1995) states that "as long as silence exists in educational data, the phenomenon of under-representation will receive critical examination, a condition that obscures the need for remedies for systemic discriminatory practices" (p. 140).

In 2021, according to the National Education Policy Center (NEPC), in order to create an environment that is equitable, welcoming and fair to everyone, systemic solutions and policies need to be implemented. "There must be accountability for equity efforts in evaluating all teachers, outlining equity-based and culturally plural norms that are used as evidence of good teaching" (NEPC, 2021, p. 1). Likewise, there must be an institutional effort to treat culturally based traits, such as the ability to speak multiple languages, as assets rather than deficits while recognizing and remunerating staff who put

them to use. “There are concrete changes that can make a district, both at the institutional/policy level and at the interpersonal/cultural level, able to substantially improve the experience of educational leaders” (NEPC, 2021, p. 3).

Recommendations for Future Research

The focus upon “raising standards in education has done very little and perhaps can do no more to close performance outcome gaps between social class groups” (Ball, 2010, p. 157). The performance gaps in terms of social class remains enormous (Lupon et al., 2021). “The community of academics, practitioners, policy-makers, and commentators that constitutes the world of Educational Leadership represents a spectrum of perspectives on what that world is like or should be like” (Gunter & Ribbins, 2003, p. 129). As Stanton et al. (2000) argue, many policy-makers may not want to know the negatives, while researchers may be overly eager to find them. Alwin (2009) asserts that leadership and its practice must respond to, as well as to take notice of, the evolving complexity and change existent issues and concerns within schools and the broader educational community.

Research must be continued in order to show the world that directly linking performance-based pay to K-12 public school teachers and other Educational Leaders does not produce the desired outcome of higher student achievement. An in-depth exploration of the IP category, which counts as one of the components of the overall teacher evaluation score, must also be investigated further in order to determine its validity and the oversight protocols that need to be in place for transparency to exist in this domain. Rockwell et al. (1998) state that “we must continue to educate diverse students in an intercultural context” (p. 3). A perpetual challenge that educators face is to

foster the development of reflection so that they will be prepared to be reflective teachers (Lyons, 2010). According to Shields (2004), through further research, a framework could be created to "alleviate the pressures of accountability, the reality of fiscal restraint, or the persistence of political interference, and it may help the educational leader to become firmly grounded in a moral and purposeful approach to leadership" (p. 110).

Researcher Reflections on the Findings of the Study

In the context of the pay-for-performance component for K-12 public school educators in this research, according to the findings of organizational psychologists, E.A. Locke et al. (1968), when subjects were paid on a piece-rate basis for their work, they found that they tended to choose easier tasks as the payment for success increased. Studies such as this have found that people working for a reward generally try to minimize challenge. Kohn (1993) supports this by stating, "It is not that human beings are naturally lazy or that it is unwise to give employees a voice in determining the standards to be used" (p. 53). Rather, people tend to lower their sights when they are encouraged to think about what they are going to get for their efforts. "Do this and you'll get that," in other words, focus attention on the "that" instead of the "this". "Emphasizing large bonuses is the last strategy we should use if we care about innovation. Do rewards motivate people? Absolutely. They motivate people to get rewards" (Kohn, 1993, p. 54). When teachers are not motivated and not appropriately paid, it could in part account for the high teacher turnover and attrition rates. Darling-Hammond (1998) stated that teacher attrition is at 30% within the first three to five years of entering the profession. "According to the numbers and predictions for enrollment, the nation will soon be in dire need of educators to teach the increasing number of students in

public schools" (p. 6). Due to the national increase of student enrollment, and the increase of teachers exiting the teaching profession, due to retirement and career changes, the dilemma at hand becomes one that affects the nation. According to Gonzalez et al. (2015), "attention has been brought to the issue of our nation's teacher shortage, but what must also be addressed and examined is the retention issue" (p. 2).

In this study, two large Florida urban based K-12 public school districts were analyzed. The input variables of FSA (R%H and M%H) and DSGs are components of the Florida Accountability Programs which consist of the FSRP and the FDAP Programs. These programs did not have statistically significant differences of the input variable means and standard deviations. They remained consistently commensurate over this 7-year longitudinal study. The question that arises is then why does such a disparity exist in the outcome variables of the percentage of teachers rated in the Highly Effective and Effective evaluation categories. As of the 2023-2024 academic school year, the FAST Progress Monitoring Testing (Florida Assessment of Student Thinking) cycle will be implemented. The three high-stakes tests are something that K-12 public school teachers will have to successfully navigate as it will also be a component of accountability in regard to merit pay salary schedule placement. There is a causal relationship with the input and output variables.

Yet, the input and output variables within these two comparable K-12 public school districts are not in alignment. The answer is not as oversimplified as just stating that one district over-pays and the other under-pays their teaching instructional faculty. There is inherent inequity embedded in the methodology that is utilized to evaluate K-12 public school teachers. Thus, the power dynamic that is a part of the IP component of the

teacher evaluation must be investigated further. The observation tool, Marzano's iObservation ®, which is utilized throughout the state is not the rudimentary cause of any bias that may occur. However, this observation tool, which is a software program, allows the observer/evaluator to input subjective qualitative information and it is transformed into a cardinal number. Depending upon the environment, school culture, and the selected principal or assistant principal who is assigned to perform the observation, the Marzano iObservation ® tool can be weaponized in the IP category of the teacher instructional evaluation to create an inaccurate rating score by the evaluator. This impacts a teacher's ability to qualify for the highest appropriate amount of base pay monetary compensation on merit pay-for-performance salary schedules.

According to a study conducted by Watlington et al. (2010), "in the Beta County School District, the annual costs associated with replacing a teacher were calculated at $12,652 per teacher" (p. 31). Moreover, the evolution from one end-of-year test into three continuous high-stakes tests supports the Barnes et al. (2007) study which "confirmed that the educational achievement of students in at-risk schools is further jeopardized by chronic teacher turnover as teachers disproportionately leave schools with high-minority, low-performing student populations" (p. 6). The K-12 public school teacher attrition has been problematic throughout the United States. However, from 2017 to 2020, the FLDOE District Reports on Instructional Staff show that there has been a decrease in Instructional Teaching Staff in the Beta School District while there has been an increase in Instructional Teaching Staff in the Pi School District (FLDOE, 2021b). Relocating to other school districts, retirement, and illness may all contribute to the aforementioned patterns. Yet, the narrative that is being presented to the public is one that is based upon a

fallacy. If teachers work hard and endeavor to provide a quality education to their students, based upon the principles of a meritocracy, they should be consistently compensated appropriately.

Therein lies the Meritocracy Paradox (Castilla & Benard, 2010). The formalized practices used to ensure meritocracy in pay have been widely expected to mitigate various forms of workplace inequality by limiting administrative use of biases rooted in non-merit factors, thereby preventing unfair treatment (Elvira & Graham, 2002; Jackson, 2010; Reskin & McBrier, 2000; Mun & Kodama, 2021). However, Castilla (2008) asserts that the failure of merit-based reward systems shows that implemented formalized performance evaluations may paradoxically increase inequality by condoning the processes that reproduce inequality and even concealing the nonmeritocratic distribution of rewards under the cover of meritocracy (Castilla & Bernard, 2010). Due to the subjectivity, bias, favoritism, and malfeasance that may occur, the accuracy of the numerical score obtained in the IP domain which is produced by the iObservation ® evaluation tool utilized in both districts can be called into question. Moreover, there are far reaching implications that could further impact policy and practice. Notwithstanding, it must be noted that the Pi School District has a more formidable merit pay salary schedule in comparison to the Beta County School District (Beta School District, 2021a; Pi District Schools Compensation Department, 2021).

Alternative theories as to reasons why a disparity existed in the outcome variables of the teacher instructional personnel evaluation ratings of K-12 public school teachers between both of the two K-12 public school districts may be due to ideological differences between the Superintendents who were in charge of each district during this

seven-year longitudinal study. Beta's Superintendent received his degrees in Economics and Business Management (Beta School District, 2021c). The Pi School District Superintendents that served during the same seven-year time period both received their Doctorate in Education (Pi District Superintendent Biography, 2021). Another rationale could be that one district tried to save money by allocating or misappropriating monetary funds to other areas. Furthermore, monetary incentives could also be attributed to these differences between the two K-12 public school districts in that the teacher instructional personnel evaluations are performed primarily by the Principals and Assistant Principals. For the sake of transparency, it is important to acknowledge that the Beta District's Administrative Staff (Principals and Assistant Principals) earn more money than their Pi School District counterparts (FLDOE, 2021b). Each K-12 public school district is responsible for training their administrations in the evaluation process. Currently, in Florida, there are no credentialing programs or a certification process to ensure that evaluators consistently and comprehensively understand how to properly perform evaluation procedures and protocols. This lack of comprehensive training could potentially cause discrepancies that serve to skew the teaching IP score between these two K-12 public school districts.

As Foucault (2000) asserts, "power is exercised only over free subjects, and only insofar as they are 'free'" (p. 342). The school is an important cultural institution in every society with a special purpose to contribute to the education of the next generation to become active, knowledgeable, and caring citizens of their societies. As a result, "the students can grow and develop into being independent and enlightened adults who are action-competent and concerned with equity. That means that the individual is able to be

a qualified participant in society" (Moos et al., 2008, p. 63). Social justice, advocacy, and activities could include such actions as "acquiring multicultural competence; working to combat racism, sexism, and homophobia; increasing access to educational and occupational opportunities; understanding and ameliorating career barriers; and empowering individuals and families" (Toporek et al., 2006, p. 10). The core skill set that is needed for social justice and advocacy work "requires the conceptual knowledge that is intimately tied to the foundation and identity of our profession" (Toporek et al., 2006, p. 14).

Consequently, considering all that K-12 public school teaching instructional faculty are held accountable for achieving, being forced onto merit pay-for-performance salary schedules does not inspire creativity and innovation as educators are not guaranteed proper placement upon the actual structure that determines their appropriate fiscal worth. Thus, educators should never have to apologize for insisting to be compensated for the worth and caliber that they bring to their profession. The United States Senate Committee on Banking, Housing, and Urban Affairs met in open session and hybrid format to conduct a hearing on January 11, 2022 regarding the nomination of the Honorable Jerome H. Powell, of Maryland, to become the Chairman of the Board of Governors of the Federal Reserve System. Senator Robert Menendez, of New Jersey, stated that "finding jobs that provide a living wage is crucial to improving our economy" (United States Senate Committee on Banking, Housing, and Urban Affairs, 2022).

It must never be forgotten that "credible and authentic Educational Leaders are expected to be transformative, to attend to social justice, and to strive for academic achievement" (Shields, 2004, p. 110). Virtual literacy, creating virtual classrooms, and

possessing technological digital proficiency are all imperative but many K-12 public school districts are still using the same antiquated methodology, protocols, evaluation procedures, and rubric in the pre and post Covid-19 world. Thus, the question may be posed as to how this research contributes to new knowledge within the teaching profession? In response, this study asserts that in order to authentically develop Transformative Educational Leaders, the antiquated form of salary schedule payment increases has to transform as well. In these modern tumultuous times that require remote teaching and learning that ranges from in person, to fully online or a hybrid teaching format, a more evolved teacher evaluation rating methodology and salary schedule are required. In every stage of our government, from the local, state, and federal levels, the professed goal is to recruit and maintain Highly Effective teachers. This research shows that if transparency, ethics, and advancements are not cultivated then the stagnation of our nation's public educational system will adversely impact the maintenance and growth of our democracy. When the next exigent instance occurs, will educators be appropriately remunerated for virtual and crisis lesson planning as well as the development and implementation of remote contextualized learning environments? In turn, successfully emerging from this global pandemic can serve as the catalyst for sustainable growth and legitimate change.

REFERENCES

Ahn, T. (2013). The missing link: Estimating the impact of incentives on teacher effort and instructional effectiveness using teacher accountability legislation data. *Journal of Human Capital, 7*(3), 230-273. https://doi.org/10.1086/673872

Alamillo, L., Palmer, D., Viramontes, C., & Garcia, E. (2005). Leaving children behind, how "Texas-style" accountability fails Latino youth. *Teachers College Record, 107*(11), 2433–2436. https://doi.org/10.1111/j.1467-9620.2005.00600.x

Alchemer, L. L. C. (2021). *About Alchemer: More ways to make your customer feedback transformative.* https://www.alchemer.com/company/about/

Alger, V. E. (2012). *Teacher selection and evaluation in Nebraska*. Independent Women's Forum. https://www.iwf.org/2012/01/11/teacher-selection-and-evaluation-in-nebraska/

Allegretto, S., & Mishel, L. (2018). *The teacher pay penalty has hit a new high: Trends in the teacher wage and compensation gaps through 2017*, 1-15. Economic Policy Institute. https://www.epi.org/publication/teacher-pay-gap-2018/

Alwin, L. (2009). *The influence of life experiences on educational leadership practice and evolution of that practice: An autoethnography*. Doctoral Dissertation - The University of Wisconsin-Madison. ProQuest LLC.

American Educational Research Association, [A. E. R. A.], American Psychological Association, [A. P. A.], & National Council on Measurement in Education, [N. C. M. E.] (2014). *Updating new educational and psychological testing standards.* PsycEXTRA Dataset. https://doi.org/10.1037/e578562014-008

American Federation for Children. (2017). *Articles, press releases, blogs, videos, contact information*. http://www.federationforchildren.org/about/press

American Federation of Teachers. (1999). *Making standards matter 1999: An update on state activity.* (Policy Brief No. 11). https://eric.ed.gov/?id=ED436607

American Institutes for Research (AIR). (2021). *Behavioral science and research organization company background information*. AIR: Mission focused, evidence driven. https://www.air.org/about-us

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education Policy Analysis Archives*, *10*(18), p. 1-74. https://doi.org/10.14507/epaa.v10n18.2002

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. Routledge.

Amrein-Beardsley, A., Collins, C., Polasky, S. A., & Sloat, E. F. (2013). Value-added model (VAM) research for educational policy: Framing the issue. *Education Policy Analysis Archives*, *21*(4), p. 1-14. https://doi.org/10.14507/epaa.v21n4.2013

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., & Wittrock, M. C. (2014). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's*. Pearson Education.

Anyon, J. (1995). Race, social class, and educational reform in an inner-city school. *Teachers College Record*, *97*(1), 69–95. https://doi.org/10.1177/016146819509700101

Apling, R. (1988). *The impact aid programs (P.L. 81-874 and P.L. 81-815): Modifications resulting from the Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988. (P.L. 100-297)*. CRS Report to Congress. https://eric.ed.gov/?id=ED300918

Apple, M. W. (2001). Markets, standards, teaching, and teacher education. *Journal of Teacher Education*, *52*(3), 182–196. https://doi.org/10.1177/0022487101052003002

Apple, M. W. (2018). Educating the right way: Markets, standards, God, and inequality. Routledge.

Arthur, G. F., & Milton, S. (1991). The Florida teacher incentive program: A policy analysis. *Educational Policy*, *5*(3), 266–278. https://doi.org/10.1177/0895904891005003004

Attanucci, J. S. (2004). Questioning honor: A parent–teacher conflict over excellence and diversity in a USA urban high school. *Journal of Moral Education*, *33*(1), 57–69. https://doi.org/10.1080/0305724042000200010

Auguste, B., Kihn, P., & Miller, M. (2010). *Closing the teaching talent gap*. McKinsey & Company. https://www.mckinsey.com/industries/education/our-insights/closing-the-teaching-talent-gap

Autor, D., Dorn, D., & Hanson, G. (2016). The China shock: Learning from labor market adjustment to large changes in trade. *Annual Review of Economics*, *8*(3), 205–240. https://doi.org/10.3386/w21906

Bacharach, S. B., Lipsky, D. B., & Shedd, J.B. (1985). *Paying for better teaching: Merit pay and its alternatives.* Organizational Analysis and Practice Publications.

Baker, E. B., & Rozendal, M. S. (2019). Cognitive-based rubrics. *Teacher Education Quarterly*, *46*(2), 58–80.

Ball, S. J. (2010). New class inequalities in education. *International Journal of Sociology and Social Policy*, *30*(3/4), 155–166. https://doi.org/10.1108/01443331011033346

Ballacci, G. (2018). *Political theory between philosophy and rhetoric: Politics as transcendence and contingency*. Palgrave Macmillan.

Ballou, D. (2001). Pay for performance in public and private schools. *Economics of Education Review*, *20*(1), 51–61. https://doi.org/10.1016/s0272-7757(99)00060-6

Ballou, D., & Podgursky, M. (1993). Teachers' attitudes toward merit pay: Examining conventional wisdom. *ILR Review*, *47*(1), 50–61. https://doi.org/10.1177/001979399304700104

Barnes, G., Crowe, E., & Schaefer, B. (2007). *The cost of teacher turnover in five school districts: A pilot study*, 1-97. National Commission on Teaching and America's Future. https://eric.ed.gov/?id=ED497176

Basileo, L. D., & Toth, M. (2019). A state level analysis of the Marzano teacher evaluation model: Predicting teacher value-added measures with observation scores. *Practical Assessment, Research, and Evaluation*, *24*(1), 1-6.

Basileo, L., Toth, M., & Kennedy, E. (2015). *Pinellas County public schools 2013-14 multiple measures pilot results*, 1-14. Learning Sciences International. http://info.learningsciences.com/rs/860TAJ757/images/PCPs%20White%20Paper%2013-14%2020150604.pdf

Benitez, M., James, M., Joshua, K., & Perfetti, L. (2017). Someone who looks like me: Promoting the success of students of color by promoting the success of faculty of color. *Liberal Education, 103*(2), 2 -8.

Benton, M., & Radziwill, N. (2016). Merit pay in higher education. *The Journal for Quality and Participation, 39*(3), 15-36.

Bergin, C., Wind, S. A., Grajeda, S., & Tsai, C. (2017). Teacher evaluation: Are principals' classroom observations accurate at the conclusion of training? *Studies in Educational Evaluation, 55*(2), 19–26. https://doi.org/10.1016/j.stueduc.2017.05.002

Berkovich, I. (2014). A socio-ecological framework of social justice leadership in education. *Journal of Educational Administration, 52*(3), 282–309. https://doi.org/10.1108/jea-12-2012-0131

Berliner, D. C. (2006). Our impoverished view of educational research. *Teachers College Record, 108*(6), 949–995. https://doi.org/10.1111/j.1467-9620.2006.00682.x

Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record, 116*(1), 1–31. https://doi.org/10.1177/016146811411600102

Beta School District [BSD]. (2021a). *School district budget 2020-2021*. https://www.betaschools.com/cms/lib/FL01803656/Centricity/Domain/13447/General%20Fund%20053120.pdf

Beta School District [BSD]. (2021b). *Beta County public schools official homepage*. Beta County Public Schools. https://www.betaschools.com/

Beta School District [BSD]. (2021c). *Superintendent's biography*. https://www.betaschools.com/Page/39239

Beta Teachers Union. (2014). *Educational professionals: Teachers contract 2013-2014 NCTQ*. https://www.nctq.org/dmsView/TeachersContract_1314_(1)

Beta Teachers Union. (2021). *Educational professionals contract: Collective bargaining agreement between - Betateachersunionononline.com*. https://www.betateachersunionononline.com/pdf/EPContract1920/Contract1922_EP.pdf

Bewley, T. (2004). *Fairness, reciprocity, and wage rigidity*. IZA Discussion papers, No. 1137. Institute for the Study of Labor (IZA). Bonn. http://hdl.handle.net/10419/20374

Bierman, H. (2011). *Private equity transforming public stock to create value*. John Wiley & Sons, Inc.

Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Measures of effective teaching*. https://files.eric.ed.gov/fulltext/ED540961.pdf

Bloom, N., & Van Reenen, J. (2007). *Measuring and explaining management practices across firms and countries* (NBER Working Paper No. 12216). National Bureau of Economic Research. https://doi.org/10.3386/w12216

Bloomquist, G. (2020). Aristotle as critic of Plato's rhetoric? Some conclusions, questions and implications. *Science Et Esprit, 72*(1-2), 49–71.

Blount, J. (2006). *Ella Flagg Young: Sage reference* (F. English, Ed.), 1085-1087. Encyclopedia of Educational Leadership and Administration. https://sk.sagepub.com/reference/edleadership/n608.xml

Blount, S. (1995). When social outcomes aren′t fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, *63*(2), 131–144. https://doi.org/10.1006/obhd.1995.1068

Bluestone, N. H. (1987). *Women and the ideal society: Plato's republic and modern myths of gender*. Berg Publishing.

Blustein, D., Elman, N., & Gerstein, L. (2001). *Executive report: Social action groups (SAGs). In report from the National Counseling Psychology conference presented to the executive boards of division* (Vol. 17). National Society of Counseling Psychology. https://doi.org/10.17744/mehc.27.3.4f18kx2gkt7905u8

Bode, B. H. (1921). Intelligence and behavior. *The Journal of Philosophy*, *18*(1), 10-17. https://doi.org/10.2307/2939295

Bogotch, I. E. (2002). Educational leadership and social justice: Practice into theory. *Journal of School Leadership*, *12*(2), 138–156.

Bogotch, I. E. (2012). Social justice in middle passage: The voyage from frustrations to hope. In M. Acker-Hocevar, J. Ballenger, A. W. Place, & G. Ivory (Eds.), *The snapshots of school leadership in the 21st century. The UCEA voices from the field project*. Information Age Publishing. https://doi.org/10.1177/105268460201200203

Bogotch, I. E., & Roy, C. B. (1997). The contexts of partial truths: An analysis of principal's discourse. *Journal of Educational Administration*, *35*(3), 234–252. https://doi.org/10.1108/09578239710170137

Borman, K. M., & Dorn, S. (2007). *Education reform in Florida: Diversity and equity in public policy*. State University of New York Press.

Borman, K. M., Eitle, T. M. N., Michael, D., Eitle, D. J., Lee, R., Johnson, L., Cobb-Roberts, D., Dorn, S., & Shircliffe, B. (2004). Accountability in a postdesegregation era: The continuing significance of racial segregation in Florida's schools. *American Educational Research Journal*, *41*(3), 605–631. https://doi.org/10.3102/00028312041003605

Bourdieu, P., & Passeron, J. C. (1990). *Reproduction in education, society, and culture*. Sage Publications. https://monoskop.org/images/8/82/Bourdieu_Pierre_Patterson_Jean_Claude_Reproduction_in_Education_Society_and_Culture_1990.pdf

Bowles, S., & Gintis, H. (2011). *Schooling in capitalist America: Educational reform and the contradictions of economic life*. Haymarket Books.

Bracey, G. W. (2008). Disastrous legacy: Aftermath of A Nation at Risk. *Dissent*, *55*(4), 80–83. https://doi.org/10.1353/dss.2008.0011

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Educational Testing Service. https://files.eric.ed.gov/fulltext/ED529977.pdf

Brehm, M., Imberman, S. A., & Lovenheim, M. F. (2017). Achievement effects of individual performance incentives in a teacher merit pay tournament. *Labour Economics*, *44*(C), 133–150. https://doi.org/10.1016/j.labeco.2016.12.008

Brewer, T. J., Myers, P. S., & Zhang, M. (2015). Islands unto themselves: How merit pay schemes may undermine positive teacher collaboration – *Critical Questions in Education, 6*(2), 45-54. (Reprinted from Smith, A. (1776). The wealth of nations. Britannica.

Brookhart, S. M. (2009). *Accountability policies and measures: What we know and what we need*. National Education Association, p. 1-7. http://www.nea.org/assets/docs/HE/AccountabilityPoliciesandMeasures.pdf

Brookings Institution. (2017). *The 2016 Education choice and competition index. With keynote remarks by U.S. Secretary of Education Betsy DeVos.* The Brookings Institution. https://www.brookings.edu/wpcontent/uploads/2017/03/20170329_ecci_devos_transcript.pdf

Brooks, J. S., Knaus, C. B., & Chang, H. (2015). Educational leadership against racism: Challenging policy, pedagogy, & practice. *International Journal of Multicultural Education, 17*(1), 1–5.

Broughman, S., Kincel, B., Willinger, K., & Peterson, J. (2021). *Characteristics of private schools in the United States: Results from the 2019-20 private school universe survey*, 1-69. National Center for Education Statistics (NCES), a part of the U.S. Department of Education. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021061

Browder, D. M., & Cooper-Duffy, K. (2003). Evidence-based practices for students with severe disabilities and the requirement for accountability in No Child Left Behind. *The Journal of Special Education, 37*(3), 157–163. https://doi.org/10.1177/00224669030370030501

Burnette, D. (2016). ESSA raises K-12 stakes in 2016 state-level elections. *Education Week, 14(6)*, 1-7. https://www.edweek.org/policy-politics/essa-raises-k-12-stakes-in-2016-state-level-elections/2016/09

Bury, J. B. (1913). *A history of freedom of thought*. Holt.

Business Wire. (2018). *Veritas Capital completes acquisition of Cambium Learning Group*. https://www.businesswire.com/news/home/20181218005160/en/Veritas-Capital-Completes-Acquisition-of-Cambium-Learning-Group

Butler, J. (2015). *Bodies that matter: On the discursive limits of sex*. Routledge.

Caballero, C., Haynes, J., & Tikly, L. (2007). Researching mixed race in education: Perceptions, policies and practices. *Race Ethnicity and Education, 10*(3), 345–362. https://doi.org/10.1080/13613320701503389

Callahan, R. E. (1964). *Education and the cult of efficiency: A study of the social forces that have shaped the administration*. University of Chicago Press.

Cambium Learning Group. (2021). *Lexia learning is the structured literacy expert*. Lexia Brands - Cambium Learning Group. https://www.cambiumlearning.com/brands/lexia

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, *2*(1), 67–90. https://doi.org/10.1016/0149-7189(79)90048-x

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305–331. https://doi.org/10.3102/01623737024004305

Carver-Thomas, D., & Darling-Hammond, L. (2017). *Teacher turnover: Why it matters and what we can do about it*. Learning Policy Institute. https://doi.org/10.54300/454.278

Castilla, E. J. (2008). Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, *113*(6), 1479–1526. https://doi.org/10.1086/588738

Castilla, E. J. (2012). Gender, race, and the new (merit-based) employment relationship. *Industrial Relations: A Journal of Economy and Society*, *51*(1), 528–562. https://doi.org/10.1111/j.1468-232x.2012.00689.x

Castilla, E. J., & Benard, S. (2010). The paradox of meritocracy in organizations. *Administrative Science Quarterly*, *55*(4), 543–676. https://doi.org/10.2189/asqu.2010.55.4.543

Center for Education Reform. (2019). *The A+ Plan – a view from the ground*. Foundation for Florida's Future. https://edreform.com/2019/05/the-a-plan-a-view-from-the-ground/

Chemerinsky, E. (2004). The deconstitutionalization of education. *Loyola University Chicago Law Journal*. 36(1), 111-135. https://www.luc.edu/media/lucedu/law/students/publications/llj/pdfs/chemerinsky.pdf

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593

Childress, H. (2019). *The adjunct underclass: How America's colleges betrayed their faculty, their students, and their mission*. The University of Chicago Press.

Chingos, M. M., & West, M. R. (2012). Do more effective teachers earn more outside the classroom? *Education Finance and Policy*, *7*(1), 8–43. https://doi.org/10.1162/edfp_a_00052

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools*. Brookings Institution Press.

Cissell, G. (2010). Kentucky and education reform: The issue of pay-for-performance. *Journal of Law and Education*, *39*(1), 119–127.

Civil Rights Act of 1964. (1964). The Civil Rights Act of 1964. *Harvard Law Review*, *78*(3), 684-696. https://doi.org/10.2307/1339128

Cochran, M., & Dean, C. (1991). Home-school relations and the empowerment process. *The Elementary School Journal*, *91*(3), 261–269. https://doi.org/10.1086/461653

Cochran-Smith, M. (2001). Constructing outcomes in teacher education. *Education Policy Analysis Archives*, *9*(11), 1-56. https://doi.org/10.14507/epaa.v9n11.2001

Cochran-Smith, M., & Fries, M. K. (2001). Sticks, stones, and ideology: The discourse of reform in teacher education. *Educational Researcher*, *30*(8), 3–15. https://doi.org/10.3102/0013189x030008003

Cohen, D. K., & Moffitt, S. L. (2010). *The ordeal of equality: Did federal regulation fix the schools?* Harvard University Press.

Cohen-Vogel, L. (2005). Federal role in teacher quality: Redefinition or policy alignment? *Educational Policy*, *19*(1), 18–43. https://doi.org/10.1177/0895904804272246

Cohen-Vogel, L., & Herrington, C. D. (2005). Introduction: Teacher and leadership preparation and development: No strangers to politics. *Educational Policy*, *19*(1), 5–17. https://doi.org/10.1177/0895904804272432

Cole, M. (2018). *Education, equality and human rights: Issues of gender, 'race', sexuality, disability and Social Class*. Routledge.

Comer, J. P. (1976). Improving the quality and continuity of relationships in two inner-city schools. *Journal of the American Academy of Child Psychiatry*, *15*(3), 535–545. https://doi.org/10.1016/s0002-7138(09)61627-7

Committee on Health, Education, Labor, and Pensions. (2010). *S. 844: Race To The Top,* 2523–2538. U.S. Dept. of Education. https://www.congress.gov/bill/111th-congress/senatebill/844?q=%7B%22search%22%3A%5B%22Senate+Bill+844%22%2C%22S.5D%7D&s=2&r=6

Compayré, G. (1971). *Jean Jacques Rousseau and education from nature*. B. Franklin.

Cooray, A., & Potrafke, N. (2011). Gender inequality in education: Political institutions or culture and religion? *European Journal of Political Economy*, *27*(2), 268–280. https://doi.org/10.1016/j.ejpoleco.2010.08.004

Corcoran, S. P. (2009). *Can teachers be evaluated by their students' test scores*? *Should they be*? Education policy for action series. [White paper]. Annenberg Institute for School Reform at Brown University. https://eric.ed.gov/?id=ED522164

Dalin, P., Kleekamp, B., & Rolff, H. (1995). *Changing the school culture*. Cassell.

Darling-Hammond, L. (1998). Teachers and teaching: Testing policy hypotheses from a national commission report. *Educational Researcher*, *27*(1), 5–15. https://doi.org/10.3102/0013189x027001005

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*(6), 1047–1085. https://doi.org/10.1111/j.1467-9620.2004.00372.x

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

Davidson, T. (2012). *Rousseau and education according to nature*. Hardpress Publishing.

Deci, E. L., & Ryan, R. M. (2010). Intrinsic motivation: *The Corsini encyclopedia of psychology*. John Wiley & Sons.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from impact. *Journal of Policy Analysis and Management*, *34*(2), 267–297. https://doi.org/10.1002/pam.21818

DeJaynes, T., Cortes, T., & Hoque, I. (2020). Participatory action research in schools: Unpacking the lived inequities of high stakes testing. *English Teaching: Practice & Critique*, *19*(3), 287–301. https://doi.org/10.1108/etpc-10-2019-0136

Dencker, J. C. (2008). Corporate restructuring and sex differences in managerial promotion. *American Sociological Review*, *73*(3), 455–476. https://doi.org/ 10.1177/000312240807300305

DeSander, M. K. (2000). Teacher evaluation and merit pay: Legal considerations, practical concerns. *Educational Assessment, Evaluation and Accountability*, *14*(4), 307-324.

Descartes, R. (1641). *Discourse on method – Third edition.* Hackett Publishing Company Inc.

Descartes, R., & Buchenau, A. (2016). *Meditations on first philosophy, 1641*. NuVision Publications.

Dewey, J. (1917). *Democracy and education: An introduction to the philosophy of education*. Macmillan Company.

Dewey, J. (1938). *Experience and education*. Macmillan Company.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record, 106*(6), 1145–1176. https://doi.org/10.1111/j.1467-9620.2004.00375.x

Diamond, S. S., & Lempert, R. O. (2018). Introduction. *Daedalus*, *147*(4), 5–14. https://doi.org/10.1162/daed_x_00516

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *The Journal of Human Resources*, *37*(4), 696-715. https://doi.org/10.2307/3069614

Duke, D. L. (2005). *Education empire: The evolution of an excellent suburban school system*. State University of New York Press.

Earley, P. M. (2000). Finding the culprit: Federal policy and teacher education. *Educational Policy*, *14*(1), 25–39. https://doi.org/10.1177/0895904800014001004

Educational Accountability Act of 1971 - [EAA]. (2021). *Educational Accountability Act of 1971: Florida Statute Ch. 71-197*. Statutes & constitution. http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&URL=0400-0499%2F0447%2F0447.html

Elementary Secondary Education Act, Public Law No. 114-95. (2015). *The Every Student Succeeds Act of 2016*. Authenticated U.S. Government Information - GPO. https://www.govinfo.gov/content/pkg/PLAW-114publ95/pdf/PLAW-114publ95.pdf

Elvira, M. M., & Graham, M. E. (2002). Not just a formality: Pay system formalization and sex-related earnings effects. *Organization Science*, *13*(6), 601–617. https://doi.org/10.1287/orsc.13.6.601.499

Emma, C., Wermund, B., & Hefling, K. (2016, December 9). Devos' Michigan schools experiment gets poor grades. *POLITICO*. https://www.politico.com/story/2016/12/betsy-devos-michigan-school-experiment-232399

English, F. (1992). History and critical issues of educational compensation systems. In L. Frase (Ed.), *Teacher compensation and motivation*. Technomic.

Every Student Succeeds Act. (2018). *Every Student Succeeds Act (ESSA)*. U.S. Department of Education. https://www.ed.gov/essa

Fairholm, M. R. (2004). A new sciences outline for leadership development. *Leadership & Organization Development Journal*, *25*(4), 369–383. https://doi.org/10.1108/01437730410538699

Ferguson, F. (1984). Reading morals: Locke and Rousseau on education and inequality. *Representations*, *6*(2), 66–84. https://doi.org/10.2307/2928538 Fernandez-

Mateo, I. (2009). Cumulative gender disadvantage in contract employment. *American Journal of Sociology*, *114*(4), 871–923. https://doi.org/10.1086/595941

Fertig, J. (2014). The iconoclastic Thomas Sowell. *Academic Questions*, *27*(2), 199–203. https://doi.org/10.1007/s12129-014-9419-2

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, *1*(5), 1–32.

Florida Contracts with Instructional Staff and Supervisors Act, Florida Stat. §1012.33 €. (2010). http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&URL=1000-1099%2F1012%2FSections%2F1012.33.html

Florida Department of Education [FLDOE]. (2004). *Merit pay for teachers: Florida legislature*. http://www.leg.state.fl.us/data/publications/2002/house/reports/EdFactSheets/fact%20sheets/meritpay.pdf

Florida Department of Education [FLDOE]. (2014). *Overview of Florida's teacher evaluation system*. https://www.fldoe.org/core/fileparse.php/7503/urlt/0102688-overviewfloridasteacherevaluationsystem.pdf

Florida Department of Education [FLDOE]. (2020a). *House Bill 641 – Funds for the operation of schools*. https://www.fldoe.org/core/fileparse.php/7749/urlt/HB641-BS.pdf

Florida Department of Education [FLDOE]. (2020b). *K-12 assessment: Computer-based testing*, 1-35. http://www.fldoe.org/accountability/assessments/k-12-student-assessment/computer-based-testing.stml

Florida Department of Education [FLDOE]. (2020c). *Teacher evaluations*. https://www.fldoe.org/newsroom/latestnews/teacherOevaluationOdataOnowOavailableOonline.stml

Florida Department of Education [FLDOE]. (2021a). *Archived statewide district evaluation results*. https://www.fldoe.org/teaching/performance-evaluation/archive.stml

Florida Department of Education [FLDOE]. (2021b). *Florida district salaries: Accountability data systems*. https://www.fldoe.org/accountability/data-sys/edu-info-accountability-services/pk-12-public-school-data-pubs-reports/staff.stml

Florida Department of Education [FLDOE]. (2021c). *Florida PK-20 education information portal: SAS® login - EdStats*. SAS® Logon Manager. https://edstats.fldoe.org/SASPortal/main.do

Florida Department of Education [FLDOE]. (2021d). *Governor DeSantis announces end of the high stakes FSA testing to become the first state in the nation to fully transition to progress monitoring*. https://www.fldoe.org/newsroom/latest-news/governor-desantis-announces-end-of-the-high-stakes-fsa-testing-to-become-the-first-state-in-the-nation-to-fully-transition-to-progress-monitoring.stml

Florida Department of Education [FLDOE]. (2021e). *Performance evaluation*. http://www.fldoe.org/teaching/performance-evaluation/

Florida Department of Education [FLDOE]. (2021f). *Progress monitoring: Building effective, data informed strategies to close achievement gaps*. https://www.fldoe.org/core/fileparse.php/19925/urlt/2-3.pdf

Florida Department of Financial Services [FLDFS]. (2021). *Florida division of accounting and auditing*. Florida Accountability Contract Tracking System. https://facts.fldfs.com/Search/ContractSearch.aspx

Florida District Court of Appeals. (2016). *Beta Teachers Union, local 1975 V. School Board of Beta County 199 so. 3D 997*. https://calculators.law/caselaw/decisions/BeZ8JqvpzRxn/beta-teachers-union-local-1975-v-school-board-of-beta-county

Florida Grading System Act, Florida Statute § 1008.34. (2021). *School report cards, district grades.* http://www.leg.state.fl.us/Statutes./index.cfm?App_mode=Display_Statute&URL=1000-1099%2F1008%2FSections%2F1008.34.html

Florida Personnel Act, Florida Statute §1012.22(1)(c)(1)(d). (2021). *Powers and duties of the district school board superintendent.* http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&URL-10001099%2F1012%2FSections%2F1012.22.html

Florida Personnel Evaluation Procedures and Criteria Act, Florida Statute §1012.34. (2020). *Statutes & constitution.* http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&URL=1000-1009%2F1012%2FSections%2F1012.34.html

Florida Senate. (2011). *Senate Bill 736: The Florida Senate*. https://www.flsenate.gov/Session/Bill/2011/0736/BillText/c1/PDF

Florida Senate. (2016). *Senate Bill 736 (2011) - The Florida Senate*. https://www.flsenate.gov/Session/Bill/2011/0736

Florida Senate. (2021). *Select committee on pandemic preparedness and response committee page: The Florida Senate*. https://www.flsenate.gov/Committees/show/PPR

Florida Standards Assessment. (2015). *Florida standards assessments 2014-15: Volume 4 evidence of reliability and validity.* Florida Department of Education. https://fsassessments.org/-/media/project/client-portals/florida/pdf/2019q1/v3fsa1516techrpt.pdf

Florida State Legislature. (2020). *Chapter 1012 section 34 - 2020 Florida Statutes: The Florida Senate*. https://www.flsenate.gov/Laws/Statutes/2020/1012.34

Florida State Legislature. (2021). *Florida school recognition program*. Statutes & constitution: http://www.leg.state.fl.us/statutes/index.cfm

Florida Student Growth Implementation Committee. (2015). *Implementation of value-added model equation formula*. Florida Department of Education. http://www.fldoe.org/teaching/performance-evaluation/student-growth-implementation-committe.stml

Foucault, M. (2000). *The subject and power*. The New Press.

Foundation for Florida's Future. (2021). *Keeping Florida first: How measuring what matters enables Florida to lead the nation in rising student achievement*, 1-5. A Florida Promise. https://afloridapromise.org/2021/09/14/keeping-florida-first-how-measuring-what-matters-enables-florida-to-lead-the-nation-in-rising-student-achievement/

Frase, L. E. (1992). *Teacher compensation and motivation*. Technomic Publishing. Company.

Fraser, N. (2000). Rethinking recognition. *New Left Review*, *3*(7), 107-120. https://newleftreview.org/issues/ii3/articles/nancy-fraser-rethinking-recognition

Fullan, M. (2003). *Change forces with a vengeance*. RoutledgeFalmer.

Fusarelli, L. D. (2004). The potential impact of the No Child Left Behind Act on equity and diversity in American education. *Educational Policy*, *18*(1), 71–94. https://doi.org/10.1177/0895904803260025

Garcia, E., & Weiss, E. (2019). *The role of early career supports, continuous professional development, and learning communities in the teacher shortage.* Economic Policy Institute. https://www.epi.org/publication/teacher-shortage-professional-development-and-learning-communities/

Garms, W. I. (1986). Merit schools for Florida. *Education and Urban Society*, *18*(3), 369–390. https://doi.org/10.1177/0013124586018003008

Gayles, J. (2007). Race, reward, and reform. *Educational Policy*, *21*(3), 439–456. https://doi.org/10.1177/0895904806289216

Gerhart, B. A., Rynes, S. L., & Rynes, S. (2003). *Compensation theory, evidence, and strategic implications*. SAGE Publications.

Gerwirtz, S., Ball, S., & Bowe, R. (1995). Market forces and parental choice. *Education Reform and Its Consequences*. Open University Press.

Gillborn, D., & Youdell, D. (2009). *Rationing education: Policy, practice, reform, and equity*. Open University Press.

Giroux, H. A. (1992). Educational leadership and the crisis of democratic government. *Educational Researcher*, *21*(4), 4–11. https://doi.org/10.3102/0013189x021004004

Goldhaber, D., Dearmond, M., & Deburgomaster, S. (2011). Teacher attitudes about compensation reform: Implications for reform implementation. *ILR Review*, *64*(3), 441–463. https://doi.org/10.1177/001979391106400302

Goldhaber, D., DeArmond, M., Player, D., & Choi, H. J. (2008). Why do so few public school districts use merit pay? *Journal of Education Finance*, *33*(3), 262–289.

Goldthorpe, J. (2003). The myth of education-based meritocracy. *New Economy*, *10*(4), 234–239. https://doi.org/10.1046/j.1468-0041.2003.00324.x

Gonzalez, L., Brown, M., & Slate, J. (2015). *Teachers who left the teaching profession: A qualitative understanding*, *13*(1), 1-13. The Qualitative Report. https://doi.org/ 10.46743/2160-3715/2008.1601

Gowlett, C., & Rasmussen, M. L. (2014). The cultural politics of Queer Theory in education research. *Discourse: Studies in the Cultural Politics of Education, 35*(3), 331–334. https://doi.org/10.1080/01596306.2014.888838

Gratz, D. B. (2010). Looming questions in performance pay. *Phi Delta Kappan, 91*(8), 16–21. https://doi.org/10.1177/003172171009100804

Green III, P. C., & Oluwole, J. O. (2015). Response to applying value-added methods to teachers in untested grades and subjects. *JL & Education*, *44*(1), 401-412.

Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal, 33*(1), 64–86. https://doi.org/10.5465/256352

Griffen, Z., & Panofsky, A. (2020). VAM on trial: Judging science in teacher evaluation lawsuits. *Journal of Cultural Economy*, *13*(4), 444–460. https://doi.org/10.1080/17530350.2020.1726792

Grogan, M., Shakeshaft, C., & Hall, B. (2011). *Women and educational leadership*. Jossey-Bass.

Gunter, H., & Ribbins, P. (2003). Challenging orthodoxy in school leadership studies: Knowers, knowing and knowledge? *School Leadership & Management, 23*(2), 129–147. https://doi.org/10.1080/1363243032000091922

Gutman, A. (1999). *Democratic education*. Princeton University Press.

Halpin, D., & Troyna, B. (1995). The politics of education policy borrowing. *Comparative Education, 31*(3), 303–310. https://doi.org/10.1080/03050069528994

Hamilton, L. S., Stecher, B., Marsh, J., McCombs, J., & Robyn, A. (2007). *Standards-based accountability under No Child Left Behind experiences of teachers and administrators in three states*. Rand Corporation.

Hammad, W. (2010). Teachers' perceptions of school culture as a barrier to shared decision-making (SDM) in Egypt's secondary schools. *Compare: A Journal of Comparative and International Education*, *40*(1), 97–110. https://doi.org/10.1080/03057920903374432

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, *8*(41), 1-26. https://doi.org/10.14507/epaa.v8n41.2000

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *The Journal of Human Resources*, *14*(3), 351. https://doi.org/10.2307/145575

Hanushek, E. A. (2002). The failure of input-based schooling policies. *The Economic* Journal. *113*(2), 64-98. *https*://doi.org/10.3386/w9040

Hargreaves, A., & Fink, D. (2004). The seven principles of sustainable leadership. *Educational Leadership*, *61*(7), 8–13.

Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Harvard Education Press.

Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education, 112*(2), 209–238. https://doi.org/10.1086/498995

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability. *American Educational Research Journal, 51*(1), 73–112. https://doi.org/10.3102/0002831213517130

Harrison, C., & Cohen-Vogel, L. (2012). The politics of teacher reform in Florida: Analyzing causal narratives surrounding state adoption of performance-based evaluations, performance pay, and tenure elimination. *Peabody Journal of Education, 87*(5), 517–534. https://doi.org/10.1080/0161956x.2012.723489

Hatry, H. P., & Greiner, J. M. (1984). *Issues in teacher incentive plans*, 285-316. National Institute of Education. https://files.eric.ed.gov/fulltext/ED244340.pdf

Hatry, H. P., Greiner, J. M., & Ashford, B. (2012). *Issues and case studies in teacher incentive plans*. Urban Institute Press.

Hawkins-Stafford Amendments. (1994). *H.R.5 - 100th congress (1987-1988): Augustus F. Hawkins & Robert T. Stafford Amendment of 1988 for Elementary and Secondary School Improvement*. United States Congressional Record. https://www.congress.gov/bill/100th-congress/house-bill/5

Hazi, H. M. (2017). VAM under scrutiny: Teacher evaluation litigation in the states. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 90*(5-6), 184–190. https://doi.org/10.1080/00098655.2017.1366803

Heneman, H. G., & Schwab, D. P. (1978). *Perspectives on personnel: Human resource management*. R.D. Irwin.

Heneman, R. L. (2002). *Strategic reward management: Design, implementations, and evaluation*. Information Age Publishing.

Herlihy, B., & Corey, G. (2014). *ACA ethical standards casebook.* American Counseling Association.

Hess, F. M. (2014). *Common sense school reform*. Palgrave Macmillan.

Hinchey, P. H. (2010). *Getting teacher assessment right: What policymakers can learn from research*, 1-34. National Education Policy Center. https://doi.org/ http://nepc.colorado.edu/publication/getting-teacher-assessment-right

Hinsdale, B. A. (1898). *Horace Mann and the Common School Revival in the United States*. C. Scribner's Sons.

Hollenbeck, S. (2018, May 16). Graduating high school in Florida will soon be harder. *ABC Action News Report*. https://www.abcactionnews.com/news/region-pinellas/graduating-high-school-in-florida-will-soon-be-harder

Honawar, V., & Olson, L. (2008). Advancing pay for performance. *Education Week, 27*(18), 26–31.

House Bill (HB) 641. (2020). *House Bill 641*. The Florida Senate. https://www.flsenate.gov/Session/Bill/2020/641

Hoxby, C. M. (2002). *The cost of accountability*, 1-27. National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w8855/w8855.pdf

Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal, 31*(5), 605–622. https://doi.org/10.1080/01411920500240767

IBM SPSS Statistics (Version 28.0) for Windows. (2021). *Version 28 IBM SPSS statistics server*. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Server_for_Windows_Installation_Instruction.pdf

Improving America's Schools Act (IASA) of 1994. (1994). *H.R.6: Improving America's Schools Act of 1994.* 103rd Congress. https://www.congress.gov/bill/103rd-congress/house-bill/6

Ingle, W. K. (2015). How changes in teacher evaluation are reallocating principals' time. School Business Affairs. *School Business Affairs, 81*(4), 15–18.

Inspector General (IG). (1996). *Semiannual report to Congress, October 1, 1995 - March 31, 1996*. United States Congress. https://doi.org/10.2172/258138

Ivashevskii, S. L. (2011). Education and ideology. *Russian Education & Society*, *53*(6), 42–48. https://doi.org/10.2753/res1060-9393530603

Jackson, R. M. (2010). *Destined for equality: The inevitable rise of women's status*. Harvard University Press.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *26*(1), 101–136. https://doi.org/10.1086/522974

Jahng, K. E. (2011). Thinking inside the box: Interrogating No Child Left Behind and Race to the Top. *KEDI Journal of Educational Policy*, *8*(1), 99–121.

Jean-Marie, G., Normore, A. H., & Brooks, J. S. (2009). Leadership for social justice: Preparing 21st century school leaders for a new social order. *Journal of Research on Leadership Education*, *4*(1), 1–31. https://doi.org/10.1177/194277510900400102

Johnson, D. (1997). Putting the cart before the horse: Parent involvement in the Improving America's Schools Act. *California Law Review, 85*(6), 1757–1795. https://doi.org/10.2307/3481082

Johnson, S. M. (1986). Incentives for teachers: What motivates, what matters. *Educational Administration Quarterly, 22*(3), 54–79. https://doi.org/10.1177/0013161x 86022003003

Joy, S., & Arellano, A. (2016). *Accountability for all: 2016: The broken promise of Michigan's charter sector - February trends in achievement*. The Education Trust - Midwest. http://edtrust.org/wp-content/uploads/2013/10/trends-in-achievement-and-attainment-since-weve-had-annual-testing-transperancy-and-serious-accountability-for-all-groups-of-children-Feb-2015.pdf

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review, 71*(4), 589–617. https://doi.org/10.1177/ 000312240607100404

Kane, T., & Staiger, D. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *National Bureau of Economic Research, 121*(5), 1–51. https://doi.org/10.3386/w14607

Kanter, R. M. (1987). *The attack on pay*. Harvard Business School Press.

Kantor, H. (2001). In retrospect: David Tyack's The One Best System. *Reviews in American History, 29*(2), 319–327. https://doi.org/10.1353/rah.2001.0033

Kaplan, L. S., & Owings, W. A. (2018). Betsy Devos's education reform agenda: What principals—and their publics—need to know. *NASSP Bulletin, 102*(1), 58–84. https://doi.org/10.1177/0192636518763084

Kasper, B. B. (2005). Educational reform 1983-1994: New ideas or the rebirth of Quintilian's ideologies? *American Educational History Journal, 32*(2), 175–182.

Kegan, R., & Lahey, L. L. (2001). *The real reason people won't change*. Harvard Business Review.

Kellett, S. (2006). *A picture of creative entrepreneurship: Visual narrative in creative enterprise education report*, 1-17. Internationalizing Entrepreneurship Education and Training Conference. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.455.7143&rep=rep1&type=pdf

Kelly, J., & Hill, D. (2002). *Marxism against Postmodernism in educational theory*. Lexington Books.

Kennedy, G. A. (1994). *A new history of classical rhetoric an extensive revision and abridgment of the art of persuasion in Greece*. Princeton University Press.

Kohn, A. (1993). Why incentive plans cannot work. *Harvard Business Review, 71*(5), 54–60.

Kothari, C. R., & Garg, G. (2019). *Research methodology: Methods and techniques*. New Age International (P) Limited, Publishers.

Kraut, R. (1997). *Plato's comparison of just and unjust lives*. Akademie Verlag.

Laker, J. A., & Davis, T. (2011). *Masculinities in higher education: Theoretical and practical considerations*. Taylor & Francis.

Lauder, H., & Hughes, D. (1999). *Trading in futures: Why markets in education don't work*. Open University Press.

Lazear, E. P. (2004). Output-based pay: Incentives, retention or sorting? *Research in Labor Economics*, *23*, 1–25. https://doi.org/10.1016/s0147-9121(04)23001-1

Leigh, A. (2013). The Economics and politics of teacher merit pay. *CESifo Economic Studies*, *59*(1), 1–33. https://doi.org/10.1093/cesifo/ifs007

Leigh, A., & Ryan, C. (2008). How and why has teacher quality changed in Australia? *Australian Economic Review*, *41*(2), 141–159. https://doi.org/10.1111/j.1467-8462.2008.00487.x

Leitch, M. L., & Tangri, S. S. (1988). Barriers to home-school collaboration. *Educational Horizons*, *66*(2), 70–74.

Lenhoff, S., Arellano, A., & Zeman, D. (2013). *Invest in what works: An education roadmap for Michigan leaders*. The Education Trust - Midwest. http://www.bridgemi.com/wp-content/uploads/2013/04/Ed-Trust-Report-4-18-13.pdf

Levy, G. (2017, January 18). Betsy Devos confirmation hearing devolves into partisan brawl on education policy. *U.S. News and World Report*. https://www.usnews.com/news/politics/articles/2017-01-18/betsy-devos-confirmation-hearing-devolves-into-partisan-brawl-on-education-policy

Lightfoot, S. L. (1981). Toward conflict and resolution: Relationships between families and schools. *Theory Into Practice*, *20*(2), 97–104. https://doi.org/10.1080/00405848109542936

Lipman, P. (2002). Making the global city, making inequality: The political economy and cultural politics of Chicago school policy. *American Educational Research Journal*, *39*(2), 379–419. https://doi.org/10.3102/00028312039002379

Lipman, P., & Gustein, E. (2001). Undermining the struggle for equity: A case study of Chicago school policy in a Latino/a school. *Race, Gender and Class*, *8*(1), 57–80.

Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness*. National Comprehensive Center for Teacher Quality. https://files.eric.ed.gov/fulltext/ED543776.pdf

Locke, J., Axtell, J., & Elton, G. R. (1968). *The educational writings of John Locke: A critical edition*. Cambridge University Press.

Loesch, D. Z. (1986). Bivariate and multivariate analysis of the skin ridge pattern intensities. *American Journal of Physical Anthropology*, *69*(3), 287–298. https://doi.org/10.1002/ajpa.1330690302

Looney, J. (2011). Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education*, *46*(4), 440–455. https://doi.org/10.1111/j.1465-3435.2011.01492.x

Lossen, U. (2007). *Portfolio strategies of private equity firms: Theory and evidence*. Deutscher Universitätsverlag.

Lunenburg, F. C., & Ornstein, A. C. (2022). *Educational administration: Concepts and practices*. SAGE Publications.

Lupon, A., Rodríguez-Lozano, P., Bartrons, M., Anadon-Rosell, A., Batalla, M., Bernal, S., Bravo, A. G., Capdevila, P., Cañedo-Argüelles, M., Catalán, N., Genua-Olmedo, A., Gutiérrez-Cánovas, C., Feio, M. J., Lucati, F., Onandia, G., Poblador, S., Rotchés-Ribalta, R., Sala-Bubaré, A., Sánchez-Montoya, M. M., Pastor, A. (2021). Towards women-inclusive ecology: Representation, behavior, and perception of women at an international conference. *PLOS ONE*, *16*(12), 1-18. https://doi.org/10.1371/journal.pone.0260163

Lyons, N. (2010). *Handbook of reflection and reflective inquiry: Mapping a way of knowing for professional reflective inquiry*. Springer Business Media.

Maher, F. A. (1999). John Dewey, progressive education and feminist pedagogies: Issues in gender, power, and authority. *Teachers College Record*, *101*(1), 35–59. https://doi.org/10.1111/0161-4681.00028

Malen, B. (2003). Tightening the grip? The impact of state activism on local school systems. *Educational Policy*, *17*(2), 195–216. https://doi.org/10.1177/0895904803017002001

Mann, H., & Cremin, L. (1979). *The republic and the school: Horace Mann on the education of free men*. Teachers College Press.

Manna, P. (2010). *Collision course: Federal education policy meets state and local realities*. CQ Press.

Manning, A., & Swaffield, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, *118*(530), 983–1024. https://doi.org/10.1111/j.1468-0297.2008.02158.x

Marsden, D., Belfield, R., Lewin, D., & Kaufman, B. E. (2007). *Advances in industrial and labor relations*. Emerald Group Publishing Limited.

Marzano Learning and Sciences Center for Teacher and Leader Evaluations. (2014). *Developing a passion for professional teaching: The Marzano teacher evaluation model*. Learning Sciences International. https://www.learningsciences.com/wp-content/uploads/2020/06/The-Marzano-Teacher-Evaluation-Model.pdf

Marzano Research. (2022). *School improvement*, 1-5. https://www.marzanoresearch.com/practice-areas/school-improvement/

Marzano, R. J. (2019). *Marzano protocol*. iObservation. https://www.iobservation.com/marzano/protocol/

Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives*. Corwin Press.

Marzano, R. J., Carbaugh, B., Rutherford, A., & Toth, M. D. (2013). *Marzano center teacher observation protocol for the 2014 Marzano teacher evaluation model*, 1-65. Learning Sciences Marzano Center Teacher and Leader Evaluation. https://sde.ok.gov/sites/default/files/documents/files/2014%20Protocol%20Paper_20140128.pdf

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2008). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Pearson.

May, L., & Delston, J. B. (2017). *Applied ethics: United Nations universal declaration of human rights*. Routledge. https://doi.org/10.4324/9781315097176-10

Mayo, M. (2012, December 12). Florida's new formula for rating teachers makes Einstein's theory look simple, 1-5. *Sun-Sentinel*. https://www.sun-sentinel.com/news/fl-xpm-2012-12-12-fl-teachers-mayocol-b121312-20121212-story.html

Mazutis, D., & Slawinski, N. (2008). Leading organizational learning through authentic dialogue. *Journal of Management Learning*, *39*(24), 437-456. https://journals.sagepub.com/doi/pdf/10.1177/1350507608093713

McGuinn, P. (2012). Stimulating reform: Race To The Top, competitive grants and the Obama education agenda. *Educational Policy*, *26*(1), 136–159. https://doi.org/10.1177/0895904811425911

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. Routledge.

Meier, S. T. (2015). *Incorporating progress monitoring and outcome assessment into counseling and psychotherapy: A* primer. Oxford University Press.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*(1), 172–175. https://doi.org/10.1037/0033-2909.111.1.172

Merit Pay Task Force. (1983). *Report to Congress: Merit pay task force conclusions*, 1-9. Unites States Congress. https://book247all.com/book/merit-pay-task-force-report/

Milstein, M. M. (1999). Reflections on the evolution of educational leadership programs. *Educational Administration Quarterly*, *35*(4), 537–545. https://doi.org/10.1177/00131619921968752

Mintrop, H. (2004). High-stakes accountability, state oversight, and educational equity. *Teachers College Record, 106*(11), 2128–2145. https://doi.org/10.1111/j.1467-9620.2004.00430.x

Mitra, A., Gupta, N., & Jenkins, Jr., G. D. (1997). A drop in the bucket: When is a pay raise a pay raise? *Journal of Organizational Behavior, 18*(2), 117–137. https://doi.org/10.1002/(sici)1099-1379(199703)18:2<117::aid-job790>3.0.co;2-1

Moos, L., Krejsler, J., Kofod, K. K. (2008). *The elusive what and the problematic how: The essential leadership questions for school leaders and educational researchers* (T. Townsend & I. Bogotch, Eds). Brill Sense Publishers.

Morreale, S. P., Valenzano, J. M., & Bauer, J. A. (2017). Why communication education is important: A third study on the centrality of the discipline's content and pedagogy. *Communication Education, 66*(4), 402–422. https://doi.org/10.1080/03634523.2016.1265136

Morris, J. D. (2022). *Morris statistical software [Computer Software] From STA 7114 statistical analysis options*. Reprinted with permission. https://johnnysolarseed.com/ContrastCorrelation.xlsx

Muijs, D., Kyriakides, L., Van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art: Teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451

Mun, E., & Kodama, N. (2021). Meritocracy at work?: Merit-based reward systems and gender wage inequality. *Social Forces, 74*(3), 1-31. https://doi.org/10.1093/sf/soab083

Munby, S. (2020). The development of school leadership practices for 21st century schools. *European Journal of Education*, *55*(2), 146–150. https://doi.org/10.1111/ejed.12394

Munby, S., & Fullan, M. (2016). *Inside-out and downside-up. How leading from the middle has the power to transform education systems*, 1-14. Education Development Trust. https://www.educationdevelopmenttrust.com/EducationDevelopmentTrust/files/51/51251173-e25d-4b34-80ae-033fcd7685ab.pdf

Murnane, R., & Cohen, D. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review*, *56*(1), 1–18. https://doi.org/10.17763/haer.56.1.l8q2334243271116

National Center for Education Statistics. (2016). *Chapter 2: Elementary and secondary education*. Digest of Education Statistics, 2014 - Elementary and secondary education. https://nces.ed.gov/programs//digest/d14/ch_2.asp

National Center for Education Statistics. (2020). *The NCES fast facts tool provides quick answers to many education questions*. NCES, a part of the U.S. Department of Education. https://nces.ed.gov/fastfacts/display.asp?id=61

National Center for Education Statistics. (2021). *Characteristics of public school teachers*. U.S. Department of Education. https://nces.ed.gov/programs/coe/indicator/clr

National Center for Fair and Open Testing. (2021). *FSA: Florida sacked the assessments*, 1-6. Florida Political Review. http://www.floridapoliticalreview.com/fsa-florida-sacked-the-assessments/

National Commission on Excellence in Education. (1983). *A Nation At Risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education.* https://eric.ed.gov/?id=ED226006

National Education Association. (2021). *Ranking of the states 2020 and estimates of school statistics. NEA research – April 2021 - nea.org*. https://www.nea.org/sites/default/files/2021/12/Immigration%20reform%20031721.pdf

National Education Policy Center. (2021). *Retaining educators of color: Some practical advice*. https://nepc.colorado.edu/sites/default/files/publications/Newsletter%20deeper%20learning.pdf

Neal, D., & Springer (ed), M. G. (2009). *Designing incentive systems for schools: Performance incentives: Their growing impact on American K-12 education*. Brookings Institution Press.

Nelson, A. R. (2016). The elementary and secondary education act at fifty: A changing federal role in American education. *History of Education Quarterly, 56*(2), 358–361. https://doi.org/10.1111/hoeq.12186

Nichols, S. L., & Berliner, D. C. (2008). *Collateral damage: How high-stakes testing corrupts America's schools*. Harvard Education Press.

Office of Program Policy Analysis and Government Accountability. (1996). *Supplementary report on the implementation and impact of Blueprint 2000 in five school districts and nineteen schools evaluated. Report No. 95-94*. The Florida Legislature. https://oppaga.fl.gov/Documents/Reports/95-54.pdf

Ong, A. (2007). Neoliberalism as a mobile technology. *Transactions of the Institute of British Geographers, 32*(1), 3–8. https://doi.org/10.1111/j.1475-5661.2007.00234.x

Oplatka, I. (2014). The place of social justice in the field of educational administration: A journals-based historical overview of emergent area of study. In I. Bogotch & C. Shields (Eds.), *International Handbook of Educational Leadership and Social (In)Justice*. Springer. https://doi.org/10.1007/978-94-007-6555-9_2

Orfield, G., & Gordon, N. (2001). *Schools more separate: Consequences of a decade of resegregation*. Civil Rights Project. Harvard University Press.

Organization for Economic Cooperation and Development. (2017). *Education at a glance: OECD indicators: United States*. OECD iLibrary. https://read.oecd-ilibrary.org/education/education-at-a-glance-2017/united-states_eag-2017-72-en

Orwin, C., & Tarcov, N. (1997). *The legacy of Rousseau*. University of Chicago Press.

Papay, J. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, *82*(1), 123–141. https://doi.org/10.17763/haer.82.1.v40p0833345w6384

Parlak, B. (2011). The relations between central administration-local administration in the states of the European Union. *Journal of TESAM Academy*, *26*(9), 7-40. https://dergipark.org.tr/en/download/article-file/137984

Pearce, J. L. (1987). Why merit pay doesn't work: Implications from organizational theory. In D. B. Balkin & L.R. Gomez-Mejia (Eds), *New Perspectives in Education* (169–178). Prentice Hall. https://escholarship.org/content/qt4fk2t3v2/qt4fk2t3v2_noSplash_b02e1729b5a28bfb3d1564dc02f4fff9.pdf

Pearce, J. L., Stevenson, W. B., & Perry, J. L. (1985). Managerial compensation based on organizational performance: A time series analysis of the effects of merit pay. *Academy of Management Journal*, *28*(2), 261–278. https://doi.org/10.5465/256200

Perez, M. C. (2011). *Inequity aversion differences: Experimental evidence among prospective teachers and lawyers*, 1-34. Center for Education Policy and Practice. https://doi.org/10.2139/ssrn.2018697

Phelps, R. P., Ferrara, L., Sadoff, R. H., & Warburton, E. C. (2005). *A guide to research in music education*. Scarecrow Press.

Pi Classroom Teachers Association. (2021). *Instructional salary information*. CTA of Pi District. https://www.picountyschools.org/UserFiles/Servers/Server_270532/File/Careers/Benefits/ChrisGraham2018.pdf

Pi District Schools Budget Department. (2021). *$3.8B budget approved for 2020-2021 Pi District school year*. School District of Pi County. https://pischools.org/news/what_s_new/september_2020/budget_approved_for_2020_2021_school_year

Pi District Schools Compensation Department. (2014). *Pi District salary schedule: Evaluation rating as of June 2014, negotiations - nctq.org*. Pi County School District. https://www.nctq.org/dmsView/Pi_District_2014-15_Salary_Schedule

Pi District Schools Compensation Department. (2021). *Pi district salary schedule: The collective bargaining - NCTQ*. Pi County School District. https://www.nctq.org/dmsView/Pi_District_CTA-July-2017-June-2020-CBA-Successor

Pi District Schools Instructional Evaluation System. (2018). *Pi District FY18-19 teacher evaluation summary instructional practice (IP)*. School District of Pi County. https://www2.pidistrictschools.org/growth/FY19TeacherEval19Summary.pdf

Pi District Superintendent Biography. (2021). *Biography of the Pi District superintendent*. School District of Pi County. https://www.pidistrictschools.org/superintendent/biography_of_the_superintendent

Pisapia, J. R. (2009). *The strategic leader: New tactics for a globalizing world*. Information Age Publishing.

Plato. (1955). *The Republic (Penguin Classics). (D. Lee, Translation). (Original work written 360 BCE).* Putnam Books.

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management, 26*(4), 909–950. https://doi.org/10.1002/pam.20292

Podgursky, M. J., & Springer, M. G. (2009). *Performance incentives: Their growing impact on American K–12 education*. Brookings Institution Press.

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, *48*(4), 965–995. https://doi.org/10.3102/0002831211410684

Preckler, M., Sainz, V., Garcia, T., Garcia, A. J., Ordoñez, E. (2012). Representations of social justice In P. Cunningham & E. Fretwell (Eds.), *Creating Communities: Local, National and Global* (538–547). Springer.

Protsik, J. (1996). History of teacher pay and incentive reforms. *Journal of School Leadership, 6*(3), 265–289. https://doi.org/10.1177/105268469600600304

Pullin, D. (2013). Legal issues in the use of student test scores and value-added models (VAM) to determine educational quality. *Education Policy Analysis Archives, 21*(4), 6-21. https://doi.org/10.14507/epaa.v21n6.2013

Raghunathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods, 1*(2), 178–183. https://doi.org/10.1037/1082-989x.1.2.178

Ramirez, A. (2001). How merit pay undermines education. *Educational Leadership, 58*(5), 16–20.

Rentner, D., & Kober, N. (2012). *After the stimulus money ends: The status of state K-12 education funding and reforms.* Center on Education Policy. https://eric.ed.gov/?id=ED529270

Reskin, B. F. (2000). The proximate causes of employment discrimination. *Contemporary Sociology, 29*(2), 319. https://doi.org/10.2307/2654387

Reskin, B. F., & McBrier, D. B. (2000). Why not ascription? Organizations' employment of male and female managers. *American Sociological Review, 65*(2), 210–233. https://doi.org/10.2307/2657438

Richardson, J. (1994). Va. district to give bonuses to top-rated teachers. *Education Week, 13(19)*, 42–51.

Ritter, G. W., & Jensen, N. C. (2010). The delicate task of developing an attractive merit pay plan for teachers. *Phi Delta Kappan, 91*(8), 32–37. https://doi.org/10.1177/003172171009100807

Rivas, A. (2015). *América Latina Después de Pisa: Lecciones aprendidas de la educación en siete países 2000-2015* [Latin America After PISA: Lessons learned from education in seven countries (2000-2015)]. Center for the Implementation of Public Policies for Equity and Growth. http://repositorio.minedu.gob.pe/handle123456789/4324

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252. https://doi.org/10.1257/0002828041302244

Rockwell, E., Anderson, G., & Montero-Sieburth, M. (1998). *Educational qualitative research in Latin America: Ethnography and the commitment to public schooling*. Garland Publishing, Inc.

Rolfe, G., Freshwater, D., & Jasper, M. (2001). *Critical reflection for nursing and the helping professions: A user's guide*. Palgrave.

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? how low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy*, *5*(2), 251–281. https://doi.org/10.1257/pol.5.2.251

Ruff, R. R. (2019). State-level autonomy in the era of accountability: A comparative analysis of Virginia and Nebraska education policy through No Child Left Behind. *Education Policy Analysis Archives, 27*(6), 1-31. https://doi.org/10.14507/epaa.27.4013

Ryu, S., & Jinnai, Y. (2020). Effects of monetary incentives on teacher turnover: A longitudinal analysis. *Public Personnel Management, 50*(2), 205–231. https://doi.org/10.1177/0091026020921414

Sanders, C. R. (2016). Money talks: The Elementary and Secondary Education Act of 1965 and the African-American freedom struggle in Mississippi. *History of Education Quarterly*, *56*(2), 361–367. https://doi.org/10.1111/hoeq.12187

Sandholtz, J. H., Ogawa, R. T., & Scribner, S. P. (2004). Standards gaps: Unintended consequences of local standards-based reform. *Teachers College Record*, *106*(6), 1177–1202. https://doi.org/10.1111/j.1467-9620.2004.00376.x

Sanford, K., Hopper, T., Robertson, K., Lancaster, L., & Collyer, V. (2019). Sustainable leadership supporting educational transformation. *In Education*, *25*(2), 3–22. https://doi.org/10.37119/ojs2019.v25i2.433

Santora, J. C. (2019). A response to global nonprofits' succession failure. *Human Resource Management International Digest*, *27*(1), 7–10. https://doi.org/10.1108/hrmid-09-2018-0181

Sawchuk, S. (2019). Florida mulls next move after governor scraps standards. *Education Week*, *38*(21), 1-7.

Scheurich, J. J., Skrla, L., & Johnson, J. F. (2000). Thinking carefully about equity and accountability. *Phi Delta Kappan*, *82*(4), 293–299. https://doi.org/10.1177/003172170008200411

Schwartz, R. E., & Fayer, H. (2006). Law, society, and democracy: Comparative perspectives. *The ANNALS of the American Academy of Political and Social Science*, *603*(1), 292–328. https://doi.org/10.1177/0002716205284197

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Scully, M. A. (2002). Confronting errors in the meritocracy. *Organization*, *9*(3), 396–401. https://doi.org/10.1177/135050840293004

Şen, S., Yıldırım, İ., & Karacabey, M. F. (2020). The implementation of the value-added assessment to determine school effectiveness. *TED EĞİTİM VE BİLİM*, *45*(202), 79–91. https://doi.org/10.15390/eb.2020.8023

Shavelson, R. J., & Marsh, H. W. (1986). *Self-related cognitions in anxiety and motivation* (R. Schwarzer, Ed.). Lawrence Eribaum Associates, Inc. Publishers.

Shelly, B. (2008). Rebels and their causes: State resistance to No Child Left Behind. *Publius: The Journal of Federalism*, *38*(3), 444–468. https://doi.org/10.1093/publius/pjn008

Shields, C. M. (2004). Dialogic leadership for social justice: Overcoming pathologies of silence. *Educational Administration Quarterly*, *40*(1), 109–132. https://doi.org/10.1177/0013161x03258963

Shinkfield, A. J., & Stufflebeam, D. (1995). *Teacher evaluation: Guide to effective practice*. Kluwer Academic.

Shulman, J. H. (1998). *Case methods in teacher education*. Teachers College Press.

Silbey, S. S. (2008). *Law and science. Epistemological, evidentiary and relational engagements*. Ashgate.

Sirotnik, K. A. (2004). *Holding accountability accountable: What ought to matter in public education*. Teachers College Press.

Skinner, B. F., Catinia (Ed), A., & Harnad, S. (1988). *The operant behaviorism of B.F. Skinner: Comments and consequences*. Cambridge University Press.

Skinner, R. R., & Kuenzi, J. J. (2015). *Reauthorization of the Elementary and Secondary Education Act: Highlights of the Every Student Succeeds Act*. Congressional Research Service. https://www.aasa.org/uploadedFiles/Policy_and_Advocacy/ESSA_Resource_Library/CRS%20ESSA%20120415.pdf

Skinner, R. R., & Rosenstiel, L. (2018). *Analysis of the Elementary and Secondary Education Act Title IA allocation formulas: Factors, design elements, and allocation patterns*. CRS Report R45141, Version 3. Updated. The Congressional Research Service. https://files.eric.ed.gov/fulltext/ED593537.pdf

Skowronek, S. (2003). *The politics presidents make: Leadership from John Adams to Bill Clinton*. Harvard University Press.

Smith, M. S., & O'Day, J. A. (1993). *Designing coherent education policy: Improving the system* (S. Fuhrman, Ed.). Jossey-Bass, Inc. Publishers.

Spivak, G. C. (2016). *Humanities, democracy, and the politics of knowledge in higher education*. Springer.

Springer, M. G. (2009). *Performance incentives: Their growing impact on American K-12 education*. Brookings Institution Press.

Stanton, G. (2000). *Further education reformed* (A. Smithers & P. Robinson, Eds.). Taylor & Francis Group.

Strain, M. M. (2005). In defense of a nation: The National Defense Education Act, Project English, and the origins of empirical research in composition. *JAC, 25*(3), 513–542.

Stucker, J. P., & Hall, G. R. (1971). *The performance contracting concept in education*. Rand Corporation. https://www.rand.org/content/dam/rand/pubs/reports/2007/R699.1.pdf

Stumbo, C., & McWalters, P. (2011). Measuring effectiveness: What will it take? *Educational Leadership, 68*(4), 10–15.

Sunderman, G. L., Kim, J. S., & Orfield, G. (2005). *NCLB meets school realities: Lessons from the field*. Corwin Press.

Terzian, S. G., & Boyd, D. C. (2004). Federal precedents and the origins of the charter school movement in Florida, USA, 1981–1996. *Journal of Educational Administration and History*, *36*(2), 135–144. https://doi.org/10.1080/0022062042000255900

Thomas, F. W., & Lang, A. R. (1937). *Principles of modern education*. Houghton Mifflin Company.

Toporek, R., Gerstein, L., Fouad, N., Roysircar, G., & Israel, T. (2006). *Handbook for social justice in counseling psychology: Leadership, vision, and action*. Sage Publications.

Torrington, D., & Weightman, J. (1989). The management of secondary schools. *Journal of Management Studies*, *26*(5), 519–530. https://doi.org/10.1111/j.1467-6486.1989.tb00742.x

Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning.* Association for Supervision and Curriculum Development.

Tyack, D. B. (1974). *The One Best System: A history of American urban education.* Harvard University Press.

Tyack, D. B., & Hansot, E. (1994). *Managers of virtue: Public school leadership in America, 1820-1980*. BasicBooks.

Tyack, D. B. & Strober, M. H. (1981). *Educational policy and management: Sex differentials* (P.A. Schmuck, W.W. Charters, & R. O. Carlson, Eds.), p. 131-152. Academic Press.

U.S. Department of Education. (2011). *Executive summary of the No Child Left Behind Act of 2001*. United States Department of Education. http://www.ed.gov/nclb/overview/intro/execsumm.html

U.S. Department of Education. (2016). *Race To The Top: Florida report*. United States Department of Education. https://files.eric.ed.gov/fulltext/ED580327.pdf

United for Florida Children. (2021). *Teacher evaluations*. UnitED For Florida Children. https://unitedforflchildren.com/research/high-stakes-testing/teacher-evaluations/

United States Senate Committee on Banking, Housing, and Urban Affairs. (2022). *Nomination hearing of the Honorable Jerome H. Powell, of Maryland, to become the chairman of the board of governors of the Federal Reserve System*. Hearings for the United States Senate Committee on Banking, Housing, and Urban Affairs. https://www.banking.senate.gov/hearings/01/04/2022/nomination-hearing

Vagi, R. (2014). Merit pay and music education: A motivation perspective. *Arts Education Policy Review*, *115*(3), 98–103. https://doi.org/10.1080/10632913.2014.914395

Vaillant, D. (2016). *Some reference frameworks in the evaluation of teacher performance*. Ibero-American Research Network on School Change and Effectiveness. https://revistas.uam.es/riee/article/download/4663/5100/0

Van Alfen, C. (1993). *School board policy as an instrument of empowering leadership in America*. Distributed by ERIC Clearinghouse.

Verger, A., Edwards, D. B., & Altinyelken, H. K. (2014). Learning from all? The World Bank, aid agencies and the construction of hegemony in education for development. *Comparative Education*, *50*(4), 381–399. https://doi.org/10.1080/03050068.2014.918713

Veritas Capital. (2020). *Leadership and investment team members*. Veritas Capital Fund Management. Retrieved from https://www.veritascapital.com/team

Warren, C. J. (1954). Brown v. Board of Education. *United States Reports*, *347*(1954), 483-492.

Watlington, E., Shockley, R., Guglielmino, P., & Felsher, R. (2010). The high cost of leaving: An analysis of the cost of teacher turnover. *Journal of Education Finance*, *36*(1), 22–37. https://doi.org/10.1353/jef.0.0028

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89–122. https://doi.org/10.3102/00346543073001089

Webb, C., Hogg, M., Borst, A., Orendorff, J. (2012). Quality communication: A 21st century leadership challenge. In G. Perreault & L. Zellner, (Eds.), *Social justice, competition and quality: 21st century leadership challenges*. National Council of Professors of Educational Administration (NCPEA). https://files.eric.ed.gov/fulltext/ED539964.pdf

Weber, C. A., & Weber, M. E. (1955). *Fundamentals of educational leadership*. Exposition Press.

Weis, L., & Fine, M. (2005). *Beyond silenced voices: Class, race, and gender in United States schools*. State University of New York Press.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project.

Weissman, S. (2021, January 4). Higher education scholars share policy hopes after Dr. Miguel Cardona's nomination for Secretary of Education, 1-5. *Diverse Issues in Higher Education*, https://www.diverseeducation.com/home/article/15108418/higher-education-scholars-share-policy-hopes-after-dr-miguel-cardonas-nomination-for-secretary-of-education

Wells, R. M. (1994). Move to revise school aid heads for troubled times. *Congressional Quarterly Weekly Report*, *52*(20), 1308–1309.

Wenning, R., Herdman, P. A., Smith, N., McMahon, N., & Washington, K. (2003). *No Child Left Behind: Testing, reporting, and accountability*. ERIC Clearinghouse on Urban Education – Development Team. https://files.eric.ed.gov/fulltext/ED480994.pdf

Wertz, S. K. (1998). Human nature and art: From Descartes and Hume to Tolstoy. *Journal of Aesthetic Education*, *32*(3), 75-89. https://doi.org/10.2307/3333307

Whitty, G. (1997). Creating quasi-markets in education: A review of recent research on parental choice and school autonomy in three countries. *Review of Research in Education*, *22*(6), 3-47. https://doi.org/10.2307/1167373

Wills, J. (2008). Subcontracted employment and its challenge to labor. *Labor Studies Journal*, *34*(4), 441–460. https://doi.org/10.1177/0160449x08324740

Wilmers, N., & Massenkoff, M. (2020). Wage stagnation and the rise of merit pay, 1974-1991. *Academy of Management Proceedings*, *2020*(1), 179-193. https://doi.org/10.5465/ambpp.2020.109

Wood, E. (1997). Modernity, postmodernity, or capitalism? *Review of International Political Economy*, *4*(3), 539–560. https://doi.org/10.14452/mr-048-03-1996-07_2

Wragg, E. C., Chamberlin, R. P., Haynes, G. S., & Wragg, C. M. (2000). *Failing teachers?*. Routledge.

Young, E. F. (2014). Isolation in the school – 1900s. *Schools: Studies in Education*, *11*(2), 321–328. https://doi.org/10.1086/678221

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*(4), 399–413. https://doi.org/10.1037/1082-989x.12.4.399