Does Rewarding Performance Pay for Teachers

Result in Higher Student Achievement?

by

Adrienne Razo

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Approved July 2014 by the
Graduate Supervisory Committee:

Dee Ann Spencer, Chair
Nicholas Appleton
Shelby Jasmer

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

In recent years, an intensifying focus has been on increasing teachers' salaries based on measurable student achievement outcomes. Funding from the Teacher Incentive Fund was followed in 2010 by Race to the Top grant, which has focused on using student achievement data to reward effective teachers. In 2010, Arizona passed legislation to adopt a model framework for a teacher evaluation instrument that included quantitative data on student achievement. School districts in Arizona needed to take a closer look at student achievement, teacher evaluation, and performance pay to determine if the models put into place accurately represent rewarding teachers for student success. This research was focused on determining the relationship between student achievement, teacher evaluation scores, and performance pay for an Arizona school district. A two-sample independent t-test was used to determine if a statistically significant difference existed in student achievement (percentages of passing and growth) on state reading scores between teachers awarded performance pay compared to teachers not awarded performance pay. In addition, correlations were conducted on overall teacher evaluation scores and each rubric within the teacher evaluation document with student reading achievement scores (percentages of passing and growth). The results of the study indicated a statistically significant difference on the percentage of students passing the state assessment for teachers who received performance pay, especially for teachers in Grades 3 through 5. Teachers' reading growth scores were not statistically significant for teachers awarded performance pay. Data analysis found no correlation between overall teacher evaluation scores in either the percentages for passing reading or for growth scores. A very weak correlation was found within the teacher evaluation rubrics of student engagement and

learning climate with AIMS reading growth scores. Further longitudinal research is recommended and changes to current teacher evaluation procedures must also be discussed.

This dissertation is dedicated to my loving *familia.* I would not be here without the enduring love of my parents Pete and Rosalie Razo. Their continued encouragement, love, and support have made me the person I am today, and I love you more than I could ever express. Whether I was frustrated during the process, needed to go to the library, or just needed help with laundry Momma was always there. These mere words on the page cannot convey my thanks enough. Thank you to all the strong women that helped raise my parents, you were an inspiration and paved the way for generations to come-thank you.

My sisters Diana, Sylvia, and Cindy, with your overwhelming love, you are a never ending source of fun. Whenever we are together, it is always entertaining, from being the loudest anywhere to just setting each other off with laughter. I love you, my sisters. Thank you for being constant cheerleaders through this process. I needed sister time then and now.

To my son Jacob, I hope you someday understand the profound change you had on my life. I would not be the person I am without you; you remain the best thing that has ever happened to me. You alone changed my life for the best in so many ways. The two of us have created an unrelenting bond and love. I am forever grateful for and, yes, you will always be my baby boy.

To the latest Razo family addition, my husband Chris. Thank you for being a distraction, a copy editor, a best friend, and a never-ending support person. You saw the good with the bad during this process, and I just want to say "thank you" for always loving and supporting me. I do need a partner, and I am glad it is you—I love you.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background

In 1983, the National Commission on Excellence in Education published *A Nation at Risk;* and for the past 30 years, America has taken a critical eye on education. Closing the achievement gap, quality teachers, and increased accountability all became public concern after *A Nation at Risk* reported American students were not able to compete globally. The public and policy makers sought increased accountability in student achievement with the goal to reduce the mediocrity of teaching declared in the report. Since the early 1900s teachers were paid based on a fixed salary schedule that awarded teachers for time on the job and additional education or higher degrees earned. In recent years, an intensifying focus has been on increasing teachers' salaries based on measurable student achievement outcomes. One of the most notable reform efforts was the No Child Left Behind Act of 2001. This act focused on efforts to attract, retain, and reward high quality teachers with the intention of increasing student achievement. No Child Left Behind defined highly qualified teachers as having the following characteristics:

1. Holding a bachelor's degree

2. Possess a full state certification or licensure

3. Being able to prove that they know each subject they teach by passing the applicable subject and/or grade-level assessment

Although No Child Left Behind has been successful in identifying teachers as highly qualified, this did not always mean these teachers were highly effective, and oftentimes,

it did not result in students attaining higher achievement scores. "How the federal government has defined highly qualified teachers, however, is somewhat limited and definitely does not capture all that it means to be an effective teacher" (Amrein-Beardsley, 2006). No Child Left Behind changed the way schools identified and selected highly qualified teachers, but the question about the relationships between teacher quality, teacher effectiveness, and student achievement still lingers.

In 2007, the Bush administration awarded the Teacher Incentive Fund (TIF Grant) to 34 states, districts, and other educational entities, with the goal of supporting projects that develop and implement performance-based compensation systems for teachers, principals, and other personnel in order to increase educator effectiveness and student achievement (Teacher Incentive Grant, 2007). Odden and Kelly (2002) defined performance pay as a "reward for specific behaviors or outcomes at the individual, team, or organizational level" (p. 59).

Using student achievement data to reward effective teachers was a cornerstone of the Obama administration's Race to the Top grant competition in 2010 (Associated Press, 2010). Since the inception of Race to the Top, 48 states have applied for a portion of the nearly four billion dollars in funding. The Center for American Progress found many states were struggling to implement their new teacher-evaluation systems, and most of the Race to the Top winners have asked to extend their timetables for completing this work (Crowe, 2012). The United States Department of Education has been tracking the performance and progress of the states that were in the first phase of applicants for Race to the Top. Although some states have been praised, others have faced distinct difficulties.

Statement of the Problem

Arizona was included in Phase 3 of states to be awarded Race to the Top grant. As part of the grant, Arizona was required to make significant changes to teacher evaluation systems. In 2010, a state law was enacted to change the culture of education in Arizona with the goal to improve how local education agencies evaluate teachers. Arizona Revised Statute § 15-203(A)(38) required the Arizona State Board of Education to adopt and maintain a model framework for a teacher evaluation instrument that included quantitative data on student academic progress. Furthermore, the statute states that student academic progress shall account for a minimum of 33% of evaluations of classroom teachers. Arizona school districts began to develop their own frameworks to comply with legislation; the state began piloting its own model in several school districts. These models were implemented in many districts in order to comply with legislation, but analysis of the effectiveness of these models is practically non-existent. School districts in Arizona need to take a closer look at student achievement, teacher evaluation, and performance pay to determine if the models Arizona has put into place accurately represent rewarding teachers for student success. This research is focused on determining the relationship between student achievement, teacher evaluation scores, and performance pay.

Purpose of Study

In 2010, Arizona mandated that school districts change their teacher evaluation methods to include student achievement data. In addition, by the 2015-2016 school year Arizona will award teachers' performance pay that is tied to teacher evaluations including

3

student achievement data. School districts around the state of Arizona have developed their own unique frameworks for the incorporation of student achievement data into teacher evaluations. These new evaluation frameworks will ultimately be tied to performance pay for teachers in Arizona; and without a careful analysis of the relationship involving student achievement, teacher evaluation scores, and performance pay, how can school districts ensure they are not awarding mediocrity?

Cardinal Valley School District (CVSD) in the central Phoenix, Arizona is an urban school district with a student enrollment of 10,118. Since 2008, the district has had a performance pay program in place tied to teacher evaluation scores and student achievement. The district performance pay plan includes the following goals:

- To make a positive impact on student achievement
- To attract and retain quality educators
- To promote cohesiveness and a cooperative spirit within the school community
- To promote individual strengths and allow for individual differences

To allow for instructional flexibility and to support instructional growth in the inception phase of the performance pay program, teacher participation was voluntary. Within the last two years, the school district of 490 teachers had 167 participants in the performance pay program and 323 teachers unable to receive performance pay due to legislative changes to the performance pay program. This created a unique situation for CVSD. Principals along with district leadership began to ask whether the teachers receiving performance pay had students with higher test scores than those teachers who did not receive the performance pay. In addition, did teachers with high student achievement scores also receive high teacher evaluation scores? As the new teacher

evaluation framework was implemented in the district, the question for the district was, Did our existing performance pay program produce higher student achievement? If our existing performance pay program was, in fact, successful in raising student achievement scores, the district's intent was to incorporate those elements into the new teacher evaluation framework. The purpose of this study was to identify whether performance pay yielded higher student achievement and determine if a relationship between student achievement and teacher evaluation scores existed within CVSD. The following research questions served as a guide for the study.

- Is there a significant difference between the proficiency levels of students on state achievement tests for teachers receiving performance pay when compared to teachers who do not receive performance pay?
- Is there a significant relationship between teacher evaluation scores and student achievement scores on state achievement tests?

Limitations

There were two limitations to the research conducted. The first limitation was the use of the state standardized achievement test for one subject during one year within CVSD. The use of one year of standardized achievement data assumes student achievement is exclusively the outcome of teacher instruction, when this is not always the case in public education. The impact of interventions, tutoring, or curriculum changes are discussed within the findings of the study. The second limitation of the study is the use of only one school district. Since only one school district was the focus of the study, the findings only represent the population of CVSD and generalized findings outside of the population in this study may not be possible.

Significance of the Study

Race to the Top was the catalyst for the new Arizona Revised Statute § 15-203(A)(38) that required the Arizona State Board of Education to implement a model framework for a teacher evaluation instrument that included quantitative data on student academic progress. Furthermore, the statute states that student academic progress shall account for at least 33% of the outcomes for teacher evaluations The Arizona State Board of Education approved the Arizona Framework for Measuring Educator Effectiveness, which complied with all legal requirements of the statute while providing school districts with some flexibility in developing their own teacher evaluation systems. As districts began the work of incorporating student achievement data into teacher evaluations and incentivizing teacher pay, the validity of all data involved had to be ensured. It is imperative for our teachers and students within Arizona that a consistent evaluation design be established to provide accurate ratings of teacher effectiveness as these ratings will determine teacher performance pay. As district and state leaders move toward performance pay for all teachers, these are all relevant questions for the CVSD superintendent and other educational leaders in Arizona.

CHAPTER 2

IMPLENTATION OF PERFORMANCE PAY PROGRAM

Chapter 2 presents the historical framework and basis for the implementation of a performance pay program to increase student achievement. The purpose of this study was to identify whether performance pay yielded higher student achievement and determine if a relationship between student achievement and teacher evaluation scores existed within Cardinal Valley School District. The first section of the literature review provides a historical perspective on education reform efforts centered on changing how teachers are compensated. The second section defines the various types of incentive-based pay and how they could be applied in an educational setting. The final section details the different performance-pay programs implemented in several states over the last decade and their outcomes to date.

Teacher Compensation

Before the discussion of why reform efforts center on changing the structure of teacher compensation, how teacher pay was established in the United States is presented. According to Spencer (2001), teaching during the 17th and 18th centuries was dominated by male clerics sanctioned by the church and only transitioned to a female-dominated profession around 150 years ago. The shift to a female workforce was due to a serious teacher shortage as the result of the beginning of the American Civil War (Clifford, 1989; Elsbree, 1939; Spencer, 2001). During the early 1800s teachers were paid on a room-and-board basis, which was successful for the rural one-room school houses of the time (Protsik, 1996). The room-and-board living arrangements also allowed for careful control and monitoring of teacher behavior, especially the behavior of unmarried young women.

Young women of the time were expected to be model citizens until they married and left the profession. "Boarding houses, teachers' clubs, and teacherages developed to provide alternate living arrangements for women teachers (Spencer, 2001 p. 807). In the late 1800s the room-and-board salary structure was replaced by paying teachers according to the grade level or subject taught. For example, a male teacher in 1876 could earn $1,700 but a woman teaching the same position could only earn $1,000. These gender gap pay differences became a focus during the women's rights movement and school districts sought a new pay structure to rectify the issue. Springer et al. (2010a), found that

> the single salary schedule determined pay according to two criteria thought to be most central to teacher productivity—years of service and degree held. It leveled the playing field relative to the grade-based compensation model by paying teachers on the same metric regardless of race, gender, or grade level taught. (p. 3)

The teacher salary schedule found in school districts today has remained largely unchanged since it was created in the early 1900s. Individual teacher attributes determine teachers' salaries based upon years of experience and education credits, so that a teacher with only two years of experience and a bachelor's degree would earn less than a teacher with four years of experience and a master's degree. The historical stepped salary schedule was created to encourage teachers to further their education while remaining employees within public education. Movement on the traditional salary schedule occurred annually for teachers as each year teachers obtained another year of experience. Given the design of the stepped salary schedule, a teacher could reach top-of-the-pay schedule after 15 years of service. Today, teachers' salaries in most of the United States' four million public schools are based on degrees and teachers' years of experience (Gonring, Teske, & Jupp, 2007). The stepped salary schedule for teachers remained an

unchanged structure and teachers have not seen significant salary increases over the last

decade.

Historical summary of public elementary and secondary school statistics: Selected years, 1919-20 through 2008-09
Data from National Center for Education Statistics http://nces.ed.gov

	1919-20	1929-30	1939-40	1949-50	1959-60	1969-70	1979-80	1989-90	1999-2000	2005-06	2006-07	2007-08	2008-09
Annual salary of classroom teachers	$9,905	$17,980	$22,348	$27,546	$36,844	$49,492	$44,585	$53,541	$53,523	$53,463	$54,202	$54,055	$54,845

Figure 1. Historical summary of public elementary and secondary school statistics

Teacher salaries began to see a slight rise in the 1980s, with the most rapid

increase between 1981 and 1990, after the appearance of *A Nation at Risk,* the national

report detailing the dire state of education in the United States (Ballou & Podgursky,

1997). *A Nation at Risk* brought many changes to education in America including the

recommendation that teachers' salaries be "professionally competitive, market sensitive

and performance based" (National Commission on Excellence in Education, 1983). Since

A Nation at Risk, major reform efforts have coupled increased student achievement with

teachers' salaries.

A History of Reform Efforts

In 1986, the Carnegie Forum on Education and the Economy created as a

response to *A Nation at Risk* a publication titled *A Nation Prepared: Teachers for the 21st*

Century. The goals of *A Nation Prepared* defined what teachers should know and be able

to do along with supporting the creation of valid assessments to certify teachers'

knowledge of their subject matter. The professionalization of teaching reform efforts

sought to align licensing teachers for what they know and what they can do for students,

while acknowledging that teaching is an intellectually complex practice (Conley &

Odden, 1995). Since *A Nation Prepared,* more national efforts followed with goals to

professionalize teaching and establish high quality standards that define effective teaching. The National Board for Professional Teaching Standards was created as a response to *A Nation Prepared,* with a purpose of developing an assessment system that would certify experienced teachers at the national level to distinguish teachers whose expertise met or exceeded high and rigorous standards (Bradley, 1994). National Board certification provided teachers a framework and guidelines as they moved from novice teachers to experienced teachers in the profession.

In 1989, President Bush convened The Education Summit composed of national political and corporate leaders, including Governor Bill Clinton. The Education Summit had six focused educational goals, including the improvement of academic performance and increased qualified teachers (Vinovskis, 1999). The Education Summit led to the establishment of the National Council on Education and Standards Testing with the goal of establishing high learning standards and assessments to ensure mastery of these standards by students by 2000. In 1994, the Clinton administration followed up the 1989 Education Summit work by passing Goals 2000, which proposed voluntary participation in testing in fourth grade reading and eighth grade math. This was surrounded by controversy and ultimately dissolved due to lack of funding. By 1997, virtually all states had teacher examinations. But simply having higher standards for teachers and better qualified teachers did not yield higher student achievement. In fact, many of these examinations were only tests of basic skills and allowed for unlimited retakes (Ballou & Podgursky, 1997). These tests of teacher quality also faced opposition from teachers and teachers' unions, who argued that a single test should not be the sole predictor of teacher effectiveness, that student success was the greatest indicator of teacher quality.

In 2001, the No Child Left Behind Act (NCLB) created the most dramatic changes to education, adding high levels of accountability by requiring both highly qualified teachers and high academic expectations of students. The belief was, in this era of high standards and high expectations, that having a highly qualified teacher had never been more important. According to NCLB, the law required all teachers of core academic subjects in the classroom be highly qualified. Teachers were highly qualified by three essential criteria: (a) attaining a bachelor's degree or better in the subject taught; (b) obtaining full state teacher certification; and (c) demonstrating knowledge in the subjects taught (*No Child Left Behind Act*, 2001). The NCLB Act established national accountability for highly qualified teachers and increased proficiency learning standards for all students without providing increased compensation or funding. With all the increased demands and accountability on teachers, their compensation structure had yet to change. In 2001, the American Federation of Teachers adopted a resolution "that supports experimentation to enhance the traditional compensation schedule using approaches that contribute to more effective teaching and learning" (Archer, 2001).

Since the inception of NCLB in 2001, states have focused on increasing their capacity to store, analyze, and report on longitudinal data for students, teachers, and schools. In light of improved data quality, some researchers and policymakers have argued that school systems should be able to estimate teachers' ability to raise student achievement and use these estimates to distinguish between more and less effective teachers. Their argument is that using these data in personnel decisions about hiring, professional development, tenure, compensation, and termination may ultimately increase the average effectiveness of the teaching workforce (Chait & Miller, 2010). Although

NCLB imposed increased accountability of all students, the goal of reaching 100%

passing for all students by 2014 was never realized.

In 2006, both government and non-profit organizations recognized the importance

of changing teacher compensation based upon teacher performance. In 2007, the Bush

administration awarded the Teacher Incentive Fund Grant (TIF) to 34 states, districts, and

other educational entities with the goal to pay teachers based on their evaluations and

their student performance scores. The TIF Grant was designed to support projects that

developed and implemented performance-based compensation systems for teachers and

principals in order to increase educator effectiveness and student achievement, as

measured by student growth, in high-need schools (US Department of Education, 2012).

In 2008, The National Center on Performance Incentives reported 26 states had at least

one initiative that tied teachers' compensation levels to their classroom performance.

State reports on the impact of performance pay on student achievement are discussed

later in this chapter.

Using student achievement growth to reward effective teachers was a cornerstone

of the Obama administration's Race to the Top grant competition in 2010 (Associated

Press, 2010). Since the inception of Race to the Top, 48 states have applied for a portion

of the nearly four billion dollars in funding. The Center for American Progress has found

> the capacity and commitment of states to implement these Race to the Top
> activities will determine success or failures and as highlighted in recent news
> reports, many states are struggling to implement their new teacher-evaluation
> systems and most of the Race to the Top winners have asked to extend their
> timetables for completing this work. (McGuinn, 2012).

Historically, reform efforts to change teacher compensation can be grouped in

three distinct incentive pay structures: career ladders, knowledge-and-skills-based, hard-

to-staff subjects, and performance pay. The next section discusses the different types of incentive pay and how they can be applied in an educational environment.

Incentive Pay

Incentive pay is defined by Merriam-Webster Dictionary as "something that incites or has a tendency to incite to determination or action" (Merriam-Webster, 2014). The basic idea behind incentive pay for employees is offering additional compensation will motivate employees to produce higher quality work or increase production levels for a reward. Incentive pay is not new to education and can be found as far back as England in the early 1700s. Gratz (2009) found "teachers' salaries in parts of the country were based on examinations of student proficiency in reading, writing, and arithmetic, though neither the schools nor this practice were systematic, and by 1890 the experiment ended.

The educational employment sector does not produce widgets; you cannot simply offer a teacher bonus pay to teach faster or teach more students. The inputs and outputs in an educational organization are more complicated than in an industrial setting. However, this did not stop continued attempts to offer incentive pay. According to Springer (2009), "Efforts to reform teacher compensation policies have emerged in virtually every decade since the 1950s. Types of reforms can be classified into a handful of categories, including performance pay, knowledge- and skills-based, career ladder programs, and hard-to-staff" (p. 4). These new compensation systems were designed to (a) promote the ongoing acquisition of skills and competencies, (b) respond to the market, (c) commit to organizational goals, and (d) accomplish results (Crandall & Wallace, 1998; Heneman, Ledford, & Gresham, 2002; Lawler, 2000; Zingheim & Schuster, 2000). The various types of incentive pay programs in education and their timeline for implementation align

with the reform efforts discussed in the first section of this chapter and are defined in

depth; the outcome of each program is discussed in the remainder of this chapter.

Career Ladders

In early 1980, as a response to *A Nation at Risk,* a Task Force on Teaching was

established with a focus on establishing an incentive pay program for teachers. Along

with establishing a new incentive pay structure for teachers, the Task Force also called

for radical changes in both teacher training and teacher certification (Hanushek, 1986).

The establishment of career ladder programs was among the first to award teachers based

on "differentiated levels of responsibility, status, and salary" (King, Swanson, &

Sweetland, 2003, p. 413). Although specific career ladder program requirements differed

by state, Gratz (2009) found program included the following goals:

- Progress on professional growth plans

- Additional training

- Differentiated duties

Career ladder programs represent job-based incentives for qualified teachers to

move into other roles in the organization. "These compensation proposals recognize the

lack of opportunity for teachers to advance professionally, except for the few who move

to administration" (Gratz, 2009, ch. 3). By recognizing and rewarding different job

functions within the schools, the requirements of the Task Force were met and for the

first time established various job classifications including master or senior teachers. For

the majority of programs, in order to be a master teacher additional trainings had to be

attended and many teachers became peer evaluators (Brandt, 1990). Career ladder

programs across the United States allowed each state and each district in those states to

determine the inputs into their career ladder plans. National studies of the effectiveness of career ladder plans are not comparable, but several states have conducted their own analysis of the impact of career ladder programs. A study of the Missouri career ladder program showed only a small positive effect on math achievement and no statistically significant impact on reading results (Glazerman &Booker, 2008). An analysis of the Arizona career ladder program found districts that participated in the program did have higher reading, math and writing scores (Amator & Kelley, 2007). National differences on the different elements of career ladder programs and the management of these programs led to funding decreases or complete elimination of career ladder programs. Modifications to previous career ladder programs helped to establish the next type of incentive pay- knowledge-and skills-based pay.

Knowledge-and-Skills-Based Pay

Although career ladder programs rewarded teachers for the additional or higher level job functions they choose, reform efforts were still focused on rewarding higher level attainment of skills and knowledge. Odden and Kelley (2002) found that "teaching to high professional standards is informed by understanding of content, knowledge about learning, and knowledge about content-specific pedagogy" (p. 19). This understanding was the basis for the NCLB demands for highly qualified teachers. The goals of the reform efforts were to assure the public school teachers where highly qualified in their subject matter and adequately prepared for the classroom environment. The National Board for Professional Teaching Standards (NBPTS) was established after the 1986 Carnegie Forum on Education report to develop an assessment system that could be used to board certify experienced teachers whose expertise met or exceeded high and rigorous

15

standards of accomplished practice. By December 2000, there were over 9,000 NBPTS

teachers across the country (Bradley, 1994; National Board for Professional Teaching

Standards, 1995; National Board for Professional Teaching Standards, 1999).

Knowledge-and skills-based pay are defined as "rewards based on completion of

teacher activities that are related to the development of knowledge and skills linked to

improved student outcomes, as well as demonstration of classroom mastery" (Springer,

2001, p. 5). Obtaining National Board Certification became an elevated position for many

teachers and school districts across the country that began to offer competitive signing

bonuses or additional stipends for teachers holding a NBPTS certification. By 2012,

rewarding NBPTS certified teachers was the most popular type of incentive pay.

As seen in Table 1, the National Center for Education Statistics found that 24.5%

of school districts rewarded NBPTS teachers some type of an incentive. Districts selected

hard-to-staff subjects as the second choice with 19.1% providing incentives to teachers

hired in these areas.

Table 1

Percentage of Public School Districts That Used Pay Incentives for Various Reasons, By Selected Public School District Characteristics (2011–12)

	To reward teachers who have attained National Board for Professional Teaching Standards certification	To reward excellence in teaching	To recruit or retain teachers to teach in a less desirable location	To recruit or retain teachers to teach in fields of shortage
All public school districts	24.5	11.3	5.6	13.5

Note. Adapted from *Schools and Staffing Survey* (SASS), by National Center for Education Statistics, 2012. Retrieved from http://nces.ed.gov/surveys/sass/tables/sass1112_2013311_d1n_005.asp)

In recognition for mastery of the rigorous NBPTS certification process, teachers obtaining the NBPTS certification were offered additional salary addenda. The website for NBPTS certification showed interested applicants how much additional addenda they could receive once they were certified. This varies by state, but at minimum, it was an additional $1,000 per year (National Board, n.d.). Some states even offered to pay fees for applicants to begin the process for National Board certification. Knowledge-and skills-based pay differed from other pay-for-performance programs because these programs "reward[ed] teachers for developing and using knowledge and skills described by external, professional standards, and identified as being valued by the school" (Odden & Kelley, 2002, p. 103). The impetus behind NBPTS is that the better qualified the teacher, the greater impact the teacher would have on student achievement.

Recent analysis of NBPTS teachers followed the implementation of the national program and several studies were initiated to answer the question. The results indicated

that there was a small positive impact on student achievement (Cavalluzzo, 2004; Clotfelter, Ladd, & Vigdor, 2007; Goldhaber, 2002; D. N. Harris & Sass, 2011). These positive gains in student achievement were enough motivation for districts to compete for NBPTS teachers; but as a result, many teachers did not seek employment in the country's neediest districts and schools. Although providing high quality teachers to all students was one of the goals of NCLB and NBPTS sought to ensure their teachers met the highest national standards, our country still struggled to place and retain these teachers for our lowest performing schools and students. The need for our best nationally recognized teachers to be compensated when accepting a teaching job in struggling schools created our next incentive pay practice of hard-to-staff subjects.

Hard-to-Staff Subjects and Schools

Hard-to-staff subjects and schools is a market-driven incentive pay program that began to evolve over the last 30 years. Hard-to-staff subjects are defined as those subjects, typically math, science, and special education, with scarcity of applicants. In addition, hard-to-staff schools are identified as typically low performing, high minority, and greater proportion of economically disadvantaged students (Ballou & Podgursky, 1997; Odden & Kelley, 2002; Springer, 2009). Along with the challenge of recruiting the most qualified teachers over the past 30 years, education has begun to face competition for teachers of math, sciences, and high technology sector jobs. Historically, teacher salaries were neither market sensitive nor market driven; by entering the teaching profession teachers did not expect to become millionaires. Now education faces fierce competition with the private sector and higher education for math and science teachers.

18

"Many districts have little difficulty in hiring elementary school teachers but face chronic shortages of applicants in special education, math, and science" (Springer, 2009).

Table 2 presents data from the National Center for Education Statistics and demonstrates the difficulty of staffing math, sciences, and special education.

Table 2

National Center for Education Statistics

	Level of difficulty filling vacancy			
Teaching field	Easy	Somewhat difficult	Very difficult	Could not fill the vacancy
General elementary	75.0	21.1	3.4	0.5
Special education	29.0	41.7	25.7	3.5
English/language arts	58.9	32.9	7.1	1.1
Social studies	71.6	24.4	3.6	0.4
Computer science	50.4	33.1	14.7	1.8
Mathematics	33.3	37.8	25.5	3.4
Biology or life sciences	34.8	44.2	19.1	1.9
Physical sciences	34.6	37.7	25.3	2.4
English as a second language	31.4	37.2	28.6	2.8
Foreign languages	26.7	39.7	29.5	4.1
Music or art	46.1	34.8	17.1	2.1
Vocational or technical education	34.4	37.7	24.2	3.7

Note. Percentages of public elementary and secondary schools with a teaching vacancy in selected teaching fields, by the school's reported level of difficulty in filling the vacancy, teaching field: 2003–04.
Adapted from *Race to the Top Program: Executive Summary* (Schools and Staffing Survey, "Public School Questionnaire," 2003–04), , by U.S. Department of Education, National Center for Education Statistics,

In a response to market demands, many districts have implemented hard-to-staff incentive pay for teachers of math, science, and special education as either a recruitment or a retention bonus (Springer & Gardner, 2010b). Many districts are faced with both hard-to-staff teaching positions within hard-to-staff schools; states have sought solutions to this problem. In 1998, Massachusetts State Legislature was facing a teacher shortage and instituted a $20,000 signing bonus to be paid over four years of consistent employment. Other places around the nation followed the trend with Houston offering

$5,000; Los Angeles with $5,000 for bilingual teachers; Dallas offered $4,500; and Nevada with a $2,000 signing bonus (Bryant, 2002; Ferdinand, 1998; Gewertz, 2001; Schemo, 2002). By offering bonuses for hard-to-staff subjects, schools assume compensation is what drives teachers to remain in their teaching positions. Liu, Johnson, and Peske (2004) found "this strategy assumes that the problem of teacher quality is getting smarter people to teach, and that the main reason smart people do not enter teaching is the low pay" (p. 222). The bonus structure sought not only to hire, but also to retain teachers in the hard-to-staff subjects and schools, so the question became, "Was this initiative successful?'

Researchers conducted a longitudinal study of the first recipients of the $20,000 recruitment bonus and found virtually no evidence that the bonus was effective at retaining teachers; teachers left for the same reasons all teachers leave the profession: site leadership, lack of support, and selection of another profession (S. Johnson & Birkeland, 2003; S. M. Johnson, Kardos, Kauffman, Liu, & Donaldson, 2004; Liu et al., 2004). With such substantial bonuses offered to teachers, the questions then became how to identify and reward quality teachers, with the anticipated outcome of increased student achievement. The desire to classify and compensate excellent teachers to effect positive student achievement results drove policy makers to the most current incentive performance-pay.

Performance Pay

After the implementation of career ladders, knowledge-skills-based, hard-to-staff subjects and schools, performance pay is the newest evolution of incentive pay in education to date. In their 2002 book, Odden & Kelley defined performance pay:

20

Performance pay represents another element of pay and compensation. Performance pay can reward specific behaviors or outcomes at the individual, team, or organizational level. At the individual level, performance pay can be called merit pay. Performance awards are usually offered as additional pay for high or improved performance. (p. 59)

Performance pay for employees in other sectors such as healthcare, business, and government are designed to accurately measure performance (output) and award accordingly (Cannon, 2007; Novicoff, 2006). Rewarding performance pay in education can be a complicated endeavor as the outputs of the educational system are not as tangible as those of the business sector. The outputs on which teachers can be assessed are varied, complicated, and unique to the grade level or subject taught. The challenges facing performance pay in education are how and which output(s) do you measure for educators and does pay for performance have a positive impact on student achievement (Kelley, Milanowski, & Heneman, 2000; Odden & Kelley, 2002; Podgursky & Springer, 2007). Performance pay output(s) for educators can be defined as follows:

- Individual teacher achievement scores on standardized tests

- Group or grade level teacher achievement scores on standardized tests

- Teacher evaluation scores (based on observational data conducted by administrators)

- School level results (or ratings) based on achievement scores on standardized tests (Ballou & Podgursky, 1997; Figlio & Kenny, 2007; Odden & Kelley, 2002; Podgursky & Springer, 2007; Springer et al., 2010a).

Colorado led the way nationally for their performance pay program which was supported by the union and their community, and became the model for national efforts funded by the United States Department of Education and private foundations.

21

Denver Professional Compensation Plan for Teachers

Denver was one of the first states to pilot and implement a performance pay program called Professional Compensation Plan for Teachers (ProComp), which linked teachers directly to their students' achievement. In 1999, the Denver School District began designing the pilot for ProComp and only one year later the performance pay program pilot testing began (Gonring, Teske, & Jupp, 2007). ProComp included components of existing incentive pay including knowledge-and-skills based along with hard-to-staff subjects and schools, but the most innovative portion of the plan was linking student achievement to the teacher and incorporating teacher evaluation as outputs for performance pay (Goldhaber & Walch, 2012; Gonring et al., 2007; Steele, Hamilton, & Stecher, 2010; Wiley, Spindler, & Subert, 2010). The pioneering ProComp model used student growth data in an attempt to accurately measure the effect quality teachers add value to a student's education. The model was created to capture the inputs into student success and then measure and reward those outputs (Goldhaber & Walch, 2012; D. Harris & Sass, 2006; Mihaly & McCaffrey, 2013; Slavin, 2008). ProComp was the first attempt to quantify the impact of a teacher during one year of instruction based upon a student's previous academic standing and is defined as "value added." "If achievement at a previous time (t^*) is also observed it is possible to concentrate on value added over the intervening period" (Hanushek, Fildes, & Davies, 1992, p. 89). For example, if a fourth grade teacher has an incoming class of students that all reached 90% proficiency on their third grade end-of-year reading assessment and leave fourth grade at only 75% proficiency, did that teacher add any value to the students in the school year?

Value added is not without controversy. Opponents argue that the variables of sorting and placement of students, student mobility, class-size, and student background all must be studied before widespread implementation (Amrein-Beardsley, 2008; Goldrick, 2002; D. Harris & Sass, 2006; Koedel & Betts, 2010; Koretz, 2002). Despite controversy with value added calculations, the ProComp pilot results were presented and approved by employees, teacher unions, and the school board in 2004, with the first checks written to teachers in 2006. Funding for ProComp has been provided by The Broad Foundation, Rose Foundation, and by Denver voters, with a $25 million dollar annual property tax increase in 2005 (Gonring et al., 2007). With significant financial, union, and public support, the next question for ProComp was "Did the program have a positive impact on student achievement?

Denver was faced with distinct groups of teachers to analyze as, during the inception, participation in ProComp was voluntary. If a teacher was employed in Denver prior to January 2006, involvement in ProComp was voluntary, whereas teachers hired after January 2006 were automatically enrolled into the performance pay program. Evaluations of ProComp found an increase in student achievement for both voluntary and non-ProComp recipients; researchers attributed this to positive changes with the entire educational system for all teachers (Goldhaber & Walch, 2012; Wiley et al., 2010). "While the findings are not completely consistent across subject and grade level, in general, we find increased achievement during the ProComp time period" (Goldhaber & Walch, 2012, p. 1077). ProComp is a successful model for the creation, implementation, and support of a performance pay program (Gratz, 2009). The United States Department of Education and private foundations watched all aspects of the ProComp implementation

23

and followed with competitive national performance pay programs, including the Teacher Incentive Fund.

Teacher Incentive Fund

Denver was one of the first successful performance pay programs able to link teacher evaluation scores and student achievement. The inclusion of hard-to-staff subjects and schools and knowledge-skills-based pay created the first comprehensive performance pay program. This comprehensive change to teacher compensation reflected the goals of multiple reform efforts since *A Nation at Risk.* In 2006, the United States Congress created a $600 million dollar federal grant program called the Teacher Incentive Fund (TIF) to support projects that reform teacher compensation. Gratz (2009) stated, "The Teacher Incentive Fund was developed as a companion to No Child Left Behind and focuses on rewarding teachers and schools for closing the achievement gap, raising student achievement, and producing real results for all children" (p. loc 3428).

The five goals of the TIF grant were similar to the elements found in ProComp:

- Improve student achievement by improving teacher and principal effectiveness

- Tie teacher and principal compensation to increases in student achievement

- Increase the number of effective teachers in hard-to-staff schools and subjects

- Create sustainable performance pay systems

- Examine multiple approaches to providing teacher incentives

Thirty three TIF grants were awarded to three state educational agencies, 22 local education agencies and eight non-profit organizations beginning in 2006 (Humphrey, Gallagher, & Yee, 2012). Figure 2 demonstrates the variety of both rural and urban districts awarded the initial TIF grants in 2006.

Notes: This exhibit shows the location of the 33 grantees in operation at the time of data collection. Urban and Rural are based on the NCES classifications in the Common Core of Data. For this report, Urban includes City and Suburb classifications; Rural includes Town and Rural classifications.

The New Leaders for New Schools grantee consisting of a consortium of charter schools implementing Effective Practice Incentive Community (EPIC) is not depicted because of its large geographic spread across many states.

In Ohio, South Dakota, and Northern New Mexico, multiple locations are implementing a statewide TIF grant. South Carolina has two TIF grants operating within the state in multiple locations.

Figure 2. Rural and urban districts awarded the initial TIF grants in 2006

Denver applied for, and was awarded, the TIF grant for their ProComp program, and in fact, 23 *of* the 33 grantees had some previous experience with various aspects of performance pay. The TIF grant allowed awardees to expand their existing program or increase payouts for teachers. Seven of the grantees also implemented the Teacher Advancement Program (TAP) created by the Milken Foundation to support their TIF grant. TAP is "a comprehensive school reform system that provides opportunities for career advancement and extensive support to teachers" (Humphrey, Gallagher, & Yee, 2012, p. 40). The first round of TIF grantees reported difficulties managing the massive amounts of student and teacher data to be created, tracked, and reported within the TIF system (McGuinn, 2012). In addition to data issues, TIF awardees found multiple problems using teacher evaluation data as part of a performance pay system.

25

TIF grant award winners had to shift their teacher evaluation from personnel driven reporting with subjective ratings to targeted instructional support and improvement systems, through numerous observations from well-trained administrators (Baker, Barton, & Darling-Hammond, 2010; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Goldrick, 2002; McGuinn, 2012; Yeh, 2009). These challenges resulted in mixed student achievement results from the first round of TIF grant award winners and spurred changes to the TIF grant structure. These included support for implementation of data systems and minimum requirements for annual teacher observations. Additional qualitative data could now be included for the performance pay including parent and student surveys of teachers (Humphrey et al., 2012; McGuinn, 2012; US Department of Education, 2012). These modifications to TIF helped shape the largest federal competitive school reform grant, Race to the Top, which was a cornerstone of Obama administration's education reform policy.

Race to the Top

In response to economic recession, President Barack Obama signed into law the American Recovery and Reinvestment Act of 2009 (ARRA). This legislation included $4.35 billion dollars in competitive grant funding, or Race to the Top grant, for educational innovation and reform. In order to apply for Race to the Top, states had to match the four core educational reform areas:

- Adopting standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy;
- Building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction;

26

- Recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most

- Turning around our lowest-achieving schools (US Department of Education, 2009, p. 2).

Many of the requirements of Race to the Top matched the refinements made to the TIF grant so it was not surprising that Colorado was among the first round of grant winners. In March 2010, the other states joining Colorado in the first round of winners were Delaware, Florida, Georgia, Illinois, Kentucky, Louisiana, Massachusetts, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, South Carolina, and Tennessee (Neil & Martinez, 2014). In addition to Race to the Top grant funding, many awardees also participated in TIF grants and Gates Foundation funded performance pay programs. States participated in multiple grants due to concern over the sustainability of funding for these various programs (Humphrey et al., 2012). In addition to sustainability concerns, Race to the Top faced criticism from teachers unions and researchers over the use of value added student growth scores and teacher evaluation scores. Despite this critique, Boser (2012), found some states appeared to be meeting the Race to the Top evaluation rubric expectations; struggling states seemed to suffer from political missteps or poor communication. In the midst of Race to the Top implementation, some states have changed academic standards and the state assessment, causing reporting on the impact of the grant to become challenging. Currently, a national report is not possible as the components selected across the states are not comparable.

Summary

This chapter has examined the establishment of teacher salaries up to the recent efforts of the Obama administration to refine performance pay for teachers. The desire to reward teachers can be found in the various incentive programs but the theme is consistent: public and policy makers have a desire to recognize and reward quality teachers who have a positive impact on student achievement. The problem education faces is the ability to accurately link performance pay, teacher evaluation, and student achievement. The question remains, "After all the changes and funds to provide performance pay for teachers, does this have a positive impact on student achievement?" The research conducted in CVSD represents an attempt to answer the question for a single school district in Arizona.

CHAPTER 3

PROCEDURES AND METHODOLOGY

This study examined the achievement levels of students whose teachers received performance pay compared to those teachers not receiving performance pay during the 2012-2013 school year. In addition, the study analyzed if a relationship existed between student achievement and teacher evaluation. A single school district was selected for the study because the components for performance pay programs and teacher evaluation instruments across Arizona are vastly different. By selecting the single school district, teachers are held to the same requirements within the performance pay program and were evaluated utilizing the same measurement instrument.

Since 2008, the CVSD performance pay program pays teachers based on their teacher evaluation scores and their student achievement outcomes. Initially, participation in the district performance pay program was voluntary for teachers. In 2012-2013, the district of 490 teachers had 167 teachers receiving performance pay, whereas the other 323 teachers were no longer eligible due to legislative changes to the performance pay program. This created a unique population for the district, principals, teachers, and the superintendent who began to ask if students of teachers receiving performance pay had higher test scores than those unable to receive performance pay. Additionally, the district wanted to determine the relationship between student achievement, teacher evaluation, and performance pay. The following questions guided this quantitative study.

Research Question 1: Is there a significant difference between the proficiency

levels of students on state achievement tests for teachers

receiving performance pay when compared to teachers who do not receive performance pay?

Research Question 2: Is there a significant relationship between teacher evaluation scores and student achievement scores on state achievement tests?

Cardinal Valley School District

CVSD is located in the Phoenix, Arizona, metropolitan area. The district operates 12 schools serving preschool through eighth grade. According to the Arizona A-F school rating system implemented in 2010-2011 by the state of Arizona, in 2012-2013, CVSD had one school rated A, eight schools rated B, three were rated with C, and the overall district grade was a B. During the 2012-2013 school year, the district had a total of 490 certified teachers with an average of 9.9 years of teaching experience.

Table 3 demonstrates the student demographic profile for CVSD, and Table 4 displays additional subgroup population data. The majority of students (70%) in CVSD are Hispanic. White students represent only 16% of the overall student population. A majority of students in CVSD (65%) qualify for free or reduced lunch and 11% of the students are English Language Learners.

Table 3

CVSD Student Demographic Data Population K-8 2012-2013

	Hispanic	White	African American	Asian	American Indian-Alaskan Native	Two or more
N = 10, 118	7,083	1,619	809	202	101	304
Percentage	70%	16%	8%	2%	1%	3%

Table 4

Additional Subgroup Student Population Data K-8 2012-2013

	Free and reduced lunch	English Language Learners	Special Education	Gifted and Talented
n = 10, 118	6,577	1,121	1,184	634
Percentage	65%	11%	12%	6%

Description of the Sample

A comprehensive list of all certified employees was provided by the Human Resources Department of the school district that included the school site, subjects and grade levels taught, and total years of teaching experience. The majority of teachers (241) in the district taught Grades 3 through 8 and administered the state assessment. For the purpose of the study, the 490 teachers were further classified and defined as described in Table 5.

Table 5

Teacher Classification

Classification	Definition	Number of teachers
Teacher K-2	Certified teacher for grades kindergarten to second grade	$n = 95$
Teacher 3-8	Certified teacher for grades third to eighth who administer the standardized state assessment	$n = 241$
Instructional Support	Certified teachers represent additional teacher leadership or instructional support for classroom teachers (for example, instructional coaches, instructional interventionist, or special education resource teachers)	$n = 108$
Special Area	Certified teachers of subjects such as physical education, arts, or music	$n = 46$
Total		$N = 490$

The sample for the research only included those teachers with the classification of

Teachers 3 through 8 for a total of 241 participants. This classification allowed the state

standardized student achievement data to be linked to the instruction of the certified

teacher in the classroom with students. Once the teacher information was collected from

human resources, student achievement results on the state standardized reading test were

obtained for Grades 3 through 8 from the district's Research and Evaluation Department.

The state standardized student achievement results were obtained for students of the

selected teachers, including their mean-scaled scores and proficiency levels. In addition

to student achievement data, teacher evaluation scores were obtained from the Human

Resources Department of the school district. Teacher evaluation scores for the district

were collected for six distinct learning targets based on a rubric designed to help teachers

improve their instructional practice. The teacher evaluation rubrics and applicable point allocations are detailed in Table 6.

Table 6

Teacher Evaluation Rubrics and Scores

Rubric name	Lesson content	Lesson facilitation	Student engagement	Learning climate	Instructional planning	Student academic progress	Total evaluation points
Points possible	5.0	5.0	5.0	5.0	6.0	5.0	31.0

Both principals and assistant principals conducted teacher evaluations two times during the school year for the school district. The final evaluation points represent the second evaluation of instructional practice. Student achievement data, teacher evaluation scores, and performance pay indication was then combined into one database for detailed analysis.

Research Design

The research was designed to determine differences between the achievement levels of students whose teachers did or did not receive performance pay, while also examining the relationship between student achievement and teacher evaluation. The groups included in the research were classroom teachers not receiving performance pay and the teachers awarded performance pay. The research design utilized was a correlation research due to the number of variables to be compared. A correlational research design is useful to researchers who are interested in determining to what degree two variables are related; however, correlational research does not "prove a relationship; rather, it indicates an association between two or more variables" (Creswell, 2008). The district

had an assumption that the teachers receiving performance pay within the district had higher levels of student achievement and higher teacher evaluation scores.

The state standardized assessment used to demonstrate reading proficiency for teachers was the AIMS (Arizona Instrument to Measure Standards). The AIMS test is a criterion-referenced assessment with embedded norm-referenced items that is administered annually in April to all students enrolled in Arizona public schools. Reading was selected as the subject for the research because all third through eighth grade teachers are assessed by the state standardized reading test. For the purpose of the study, reading results for the 2012-2013 school year were evaluated in terms of reaching proficiency and differences in scale score results by May 2013. Reading proficiency for the research was defined as a student reaching passing on the assessment administered. Each grade level has different scales for their reading assessment, but each of these scales has a cut point to indicate passing the assessment. For this research, any student reaching the passing point for his or her appropriate grade level assessment was considered proficient on the reading assessment. Math was not selected for analysis during the 2012-2013 school year as CVSD implemented a new curriculum and this would have had an impact on student scores.

<div align="center">**Variables**</div>

For this study, the independent variable was if teachers were or were not awarded performance pay during the 2012-2013 school year. The dependent variables for the study were the mean of students reaching reading proficiency on state assessments and individual teacher evaluation scores. The list of classroom teachers was labeled as either having received performance pay or not receiving performance pay. As seen in Table 7,

for 2012-2013, 153 classroom teachers did not receive performance pay and 88

classroom teachers were awarded performance pay.

Table 7

Classification of Teachers and Performance Pay

Grade	No Awarded Performance Pay	Yes Awarded Performance Pay	Total
3rd	19	17	36
4th	26	13	39
5th	27	12	39
6th	22	16	38
7th	33	16	49
8th	26	14	40
Total teachers: *Grades 3-8*	*153*	*88*	*241*

The average years of teaching experience for classroom teachers awarded incentive pay

was 10.9 with the average of teachers not receiving incentive pay was 8.5.

Data Analysis

The data analysis utilized for the study was the two-sample independent *t* test.

Glass and Hopkins (1970) found *t* test useful when comparing the mean differences

between independent groups and allowing justification for the conclusion that the

difference did not occur by random chance alone. The purpose of the two-sample *t* test

was to understand if there was an interaction between the independent variables (awarded

performance pay "yes" or "no") on the dependent variables (student achievement scores

and teacher evaluation scores) for this research study. The following questions guided the

data analysis of the research:

Research Question 1: Is there a significant difference between the proficiency

levels of students on state achievement tests for teachers

receiving performance pay when compared to teachers who

do not receive performance pay?

Research Question 2: Is there a significant relationship between teacher

evaluation scores and student achievement scores on state

achievement tests?

Assumptions

Each statistical test performed has a set of assumptions that needs to be outlined

before any analysis can be conducted. The *t* test assumes the samples have close or equal

variance. To determine if the variances between the two samples are equal, Glass and

Hopkins (1970) demonstrated the use of the Levene Test for homogeneity of variances

that will determine if the populations have equal or unequal variances. "If a random

sample of persons receives a special treatment and a second independent sample does not,

the two results means are said to be independent" (p. 284). If the samples have equal

variances, the *t* test will justify the conclusion that the difference did not occur by random

chance alone.

Instruments

The research included the reading achievement data for teachers with students in

third through eighth grade during the 2012-2013 school year in CVSD. The reading

assessment utilized for the teachers was the AIMS reading test for students in third

through eighth grade. In 2012-2013, AIMS was a dual purpose assessment containing

both criterion-referenced and norm-referenced test items designed to assess student

mastery of Arizona state instructional standards. AIMS is administered annually in April to all students in Grades 3 through 8 in reading and math. The AIMS test is used as high school exit criteria; all students in Arizona in 10th grade must pass the reading and math portions to graduate. The AIMS test is also utilized for state accountability purposes including the labeling of schools in the new A-F school accountability model. AIMS reading and math were administered during April of 2012-2013 to all students in Grades 3 through 8 enrolled in CVSD. The Arizona standardized AIMS test categorizes students within four outcomes in reading for Grades 3 through 8. The Arizona Department of Education defines outcomes on the AIMS Reading assessment using the following labels:

- Exceeds the Standard: Students who score at this level illustrate a superior academic performance as evidenced by performing substantially beyond the achievement goal for all students.

- Meets the Standard: Students who score at this level demonstrate a solid academic performance on subject matter.

- Approaches the Standard: Students who score at this level show partial understanding of the knowledge and application of the skills that are fundamental for proficient work.

- Falls Far Below the Standard: Students who score at this level may have significant gains and limited knowledge and skills that are necessary to satisfactorily meet the state's reading standard.

Students categorized as Exceeds the Standard and Meets the Standard are considered as having passed the AIMS test. The AIMS reading test has clear definitions of reading proficiency for students and is administered to all students in the state of

Arizona, and for these reasons, it was selected for use in this study. The AIMS math test was not considered for part of the research because the math standards had recently changed and instruction in CVSD for math was on the new standards. In 2012-2013, an alignment in math instruction and the state math assessment was not possible and therefore the data was not included.

A committee of teachers, district, and school site administrators with the help of a national consulting firm designed the teacher evaluation document utilized by CVSD. The teacher evaluation document has remained unchanged since it was created in the year 2005-2006. Annually CVSD reviews teacher evaluation data and professional development is provided for all leadership specifically for inter-rater reliability. Annually each administrator must attend Qualified Evaluator training to ensure understanding and consistent use of the teacher evaluation documentation.

This chapter detailed the sample and methods used for this quantitative study analyzing the proficiency levels of students on achievement tests for teachers receiving performance pay when compared to teachers who do not receive performance pay. The following chapter presents the results of the analysis of the data.

CHAPTER 4

FINDINGS

This chapter reports the findings of this study that examined the achievement levels of students whose teachers received performance pay compared to those teachers not receiving performance pay during the 2012-2013 school year. In addition, the study analyzed data to determine if a relationship existed between student achievement results and teacher evaluation scores. The following research questions guided this investigation.

Research Question 1: Is there a significant difference between the proficiency levels of students on state achievement tests for teachers receiving performance pay when compared to teachers who do not receive performance pay?

Research Question 2: Is there a significant relationship between teacher evaluation scores and student achievement scores on state achievement tests?

Sample and Data

For 2012-2013, 153 classroom teachers did not receive performance pay and 88 classroom teachers had been awarded performance pay. Of these teachers, 115 represented teaching third through fifth grade, and 126 taught sixth through eighth grade for a total of 241 teachers included in the analysis (see Table 8).

Table 8

Grade Level Groups of Teachers and Performance Pay

Grade	No Performance Pay	Yes Performance Pay	Total
3rd-5th	73	42	115
6th -8th	80	46	126
Total teachers: Grades 3-8	*153*	*88*	*241*

Teacher evaluation scores and student achievement scores in reading on the Arizona Instrument to Measure Standards (AIMS) state assessment were collected and analyzed for each of the 241 teachers included in the study. Data analysis was conducted on reading achievement student scores, including the percentage for passing AIMS reading, reading growth, and teacher evaluation scores utilizing the Statistical Program for the Social Sciences (SPSS). Analysis included correlation, Levene's test, and *t* tests to determine the statistical significance of the findings.

Results

Research Question 1

An independent samples *t* test was conducted to compare the achievement scores for teachers awarded performance pay with those teachers not awarded performance pay. The independent variable was performance pay "yes" or "no" and the dependent variables were AIMS reading achievement scores and teacher evaluation scores. The first research question asked, *Is there a significant difference between the proficiency levels of students on state achievement tests for teachers receiving performance pay when compared to teachers who do not receive performance pay?*

A *t* test analysis was conducted for teachers awarded performance pay, "yes" or "no," along with the percentage passing the AIMS reading test. The results were $t(234) = 2.92$, $p = .004$ (two-tailed) in scores for Performance Pay "Yes" ($M = 80.1$, $SD = 10.1$) and Performance Pay "No" ($M = 75.3$, $SD = 15.3$), with a mean difference of 4.78 and a 95% confidence interval of 1.19 to 8.37. The Levene's test of .06 signifies equal variances and based on the $p = .004$, there was a statistically significant difference between the percentages of the passing scores for AIMS reading of those teachers who did receive performance pay when compared with those teachers who did not receive performance pay. Table 9 displays the mean percentage of students passing the AIMS reading test for teachers awarded performance pay compared to teachers not receiving performance pay, followed by Figure 3, which is a graphic representation of the data.

Table 9

Percentage of Means for Passing AIMS Reading Comparison

	Count	Percentage of means passing AIMS Reading	Standard Deviation
Performance Pay Yes	88	80.1	10.1
Performance Pay No	153	75.3	15.3

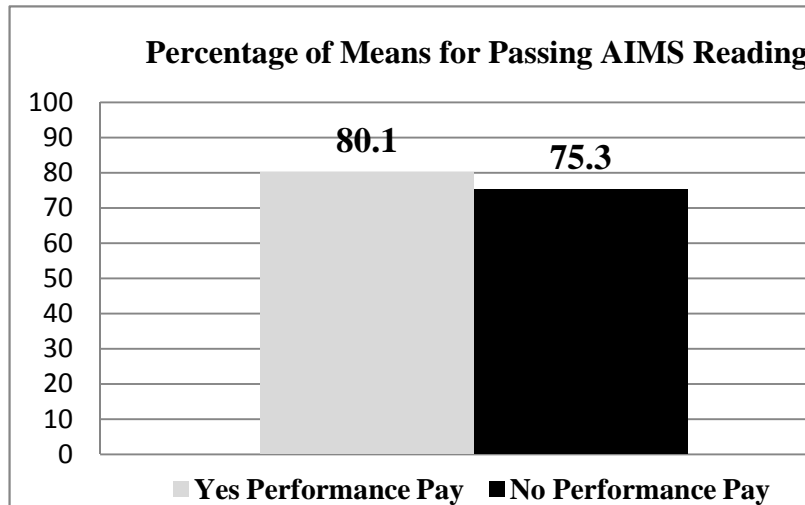**Percentage of Means for Passing AIMS Reading**

Figure 3. Percentage of means for passing AIMS reading tests by performance pay categories

An additional *t* test was conducted to determine if there was a statistically significant difference in the AIMS reading growth scores. AIMS growth is calculated annually by the Arizona Department of Education (ADE), Research and Evaluation Department. Growth as calculated by ADE is modeled after the Colorado growth model and the research conducted by Damian Betebenner (2011) who stated,

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

Following the research conducted by Damian Betebenner, the ADE technical manual annually calculates student growth scores for reading and math using quantile regression to establish curvi-liner functional relationships between students' prior year and current year scores (Huppenthal, 2013). This calculation allows school districts another measure of student achievement along with measuring the amount of students passing the state assessment.

Teachers receiving performance pay were then compared with those teachers not receiving performance pay to determine if a difference existed between student growth scores. An independent *t* test failed to reveal a statistically reliable difference between the AIMS reading growth scores for teachers performance pay "yes" (*M* = 52.9, *SD* = 9.1) and performance pay "no" (*M* = 51.3, *SD* = 9.2), a mean difference of 1.58, $t(239) = 1.29$, $p = .200$ (two-tailed). The Levene's test result of .72 represents equal variances; and with a $p = .20$; there was not a statistically significant difference between the AIMS reading growth scores for teachers awarded performance pay when compared to those not awarded performance pay. Table 10 displays the number of teachers in each category of performance pay, either "yes" or "no," and their mean AIMS growth score and the standard deviation. Figure 4 is a graphical representation of each teacher population and their mean AIMS reading growth score.

Table 10

Means for AIMS Reading Growth Score Comparison

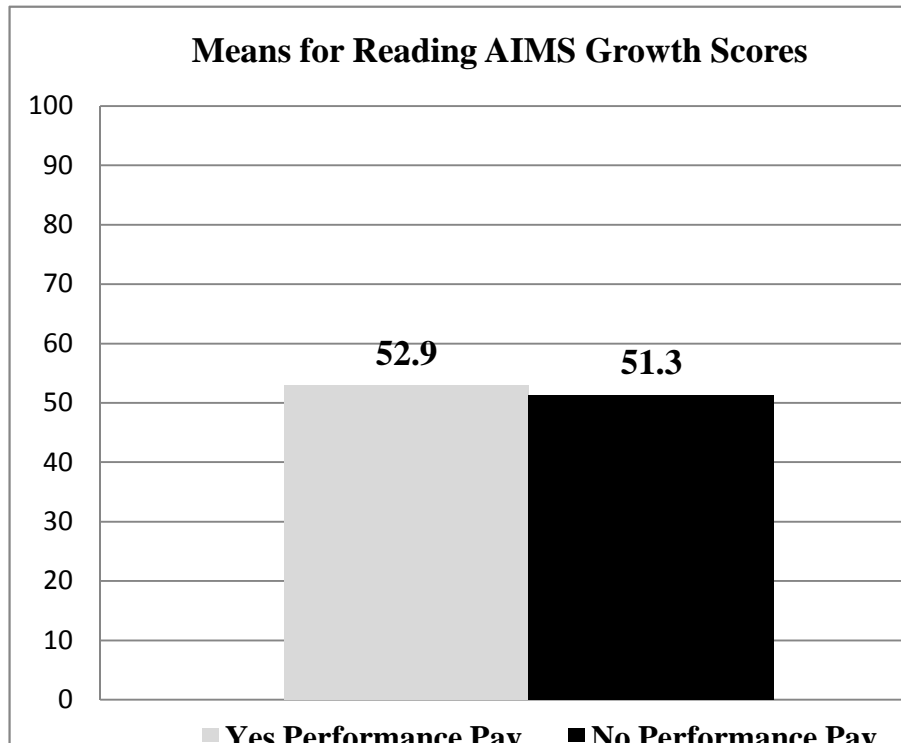|	Count	Means for AIMS reading growth scores	Standard deviation
Performance Pay Yes	88	52.9	9.1
Performance Pay No	153	51.3	9.2

Means for Reading AIMS Growth Scores

Figure 4. Means for AIMS reading growth scores by performance pay categories

Further data analysis was conducted on the teacher groups. The results for the first group of teachers, those who taught third through fifth grades, and either did or did not receive performance pay are reported below. The results for the third through the fifth group with scores for the third through fifth grade teachers receiving performance pay demonstrated a statistically significant difference between the percentage passing AIMS reading for teachers in Grades 3 through 5 who received performance pay ($M = 78.2$, $SD = 10.8$) and the third through fifth grade teachers not receiving performance pay ($M = 70.1$, $SD = 14.8$), a mean difference of 4.78, $t(106) = 3.36$, $p = .001$ (two-tailed). The same analysis was conducted on the group of teachers who taught Grades 6 through 8 and either were or were not awarded performance pay. The $t(122)$ $n = .83$, $p = .406$ (two-tailed) with scores for sixth through eighth grade teachers awarded performance pay ($M =$

81.7, *SD* = 9.1) and those teachers of sixth through eighth grades not awarded

performance pay (*M* = 80.0, *SD* =14.1) showed a mean difference of 1.7 with a 95%

confidence interval of –2.37 to 5.84. A Levene's test result of .51 assumes variances are

equal. A *p* = .406 showed there was not a statistically significant difference between the

percentages for passing scores in reading for teachers in Grades 6 through 8 who either

received performance pay or did not receive performance pay. Table 11 displays the

number of teachers in each grade level grouping, their mean AIMS reading percentages

for the AIMS reading passing scores, and the standard deviation. Figure 5 visually

displays the difference in the percentages of the means for the AIMS reading scores for

both third through fifth and the sixth through eighth grade level groups.

Table 11

Percentage of Means for Passing AIMS Reading by Grade Level Groups Comparison

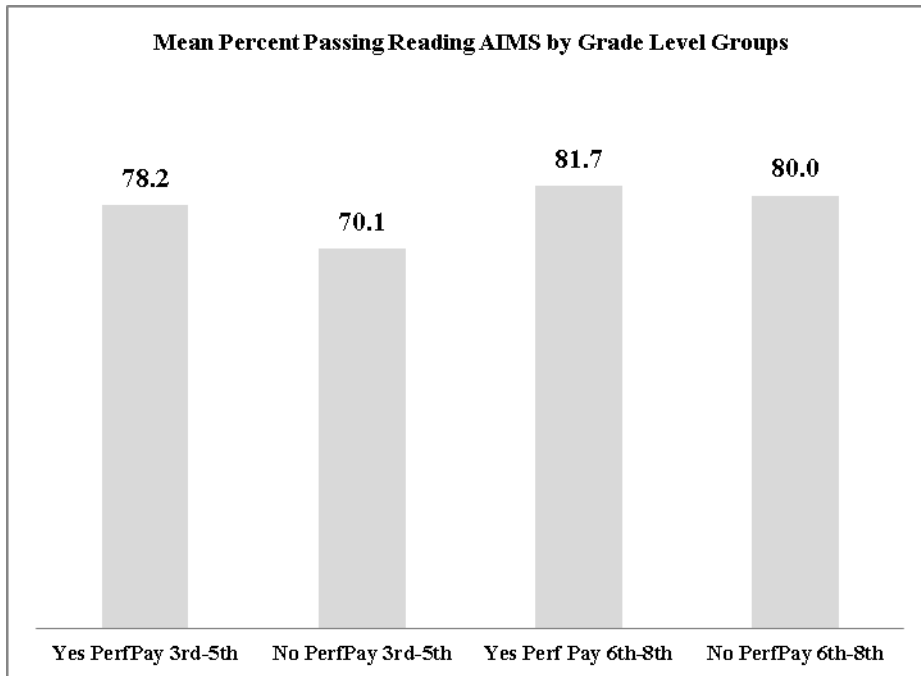|	Count	Percentage of means passing AIMS reading	Standard deviation
3rd-5th Performance Pay Yes	42	78.2	10.8
3rd-5th Performance Pay No	73	70.1	14.8
6th-8th Performance Pay Yes	46	81.7	9.1
6th-8th Performance Pay No	80	80.0	14.1

Figure 5. Percentage of means for passing AIMS reading by grade-level groups

Further analysis was conducted on each of the grade level groups to determine if there was a statistically significant difference between the grade level groups and the AIMS reading growth scores for the teachers in the different performance pay categories. For teachers in third through fifth grade, when comparing their AIMS reading growth scores, a statistically significant difference was not demonstrated. Teachers in Grades 3 through 5 awarded performance pay had AIMS reading growth scores of ($M = 50.5$, $SD = 10.5$); whereas, teachers in Grades 3 through 5 not awarded performance pay had AIMS reading growth scores of ($M = 46.9$, $SD = 9.5$). This represented a mean difference of 3.60 and a 95% confidence interval of $-.212$ to 7.42, $t(113) = 1.87$, $p = .064$ (two-tailed). The Levene's test result of .326 with $p = .064$ indicated there was not a statistically significant difference between the AIMS reading growth scores for teachers in Grades 3

through 5 who were either awarded performance pay or not during the 2012-2013 school year.

The second grade level grouping of sixth through eighth grade was also compared by both performance pay categories and their AIMS reading growth achievement data. AIMS reading growth scores for sixth through eighth grade performance pay recipients ($M = 55.1$, $SD = 6.9$) and sixth through eighth grade not performance pay recipients ($M = 55.3$, $SD = 6.8$) a mean difference of –.26 with a confidence interval of –2.78 to 2.26, $t(124)= -.2$, $p = .838$(two-tailed) Levene's test resulted in .939. This represented that a statistically significant difference did not exist in the AIMS reading growth scores for teachers in Grades 6 through 8 in either performance pay category. Table 12 displays the grade level group count of teachers, mean AIMS reading growth score, and standard deviation, followed by Figure 6 which displays the mean AIMS reading growth score for each of the grade levels and performance pay categories.

Table 12

Mean AIMS Reading Growth Score by Grade Level Groups Comparison

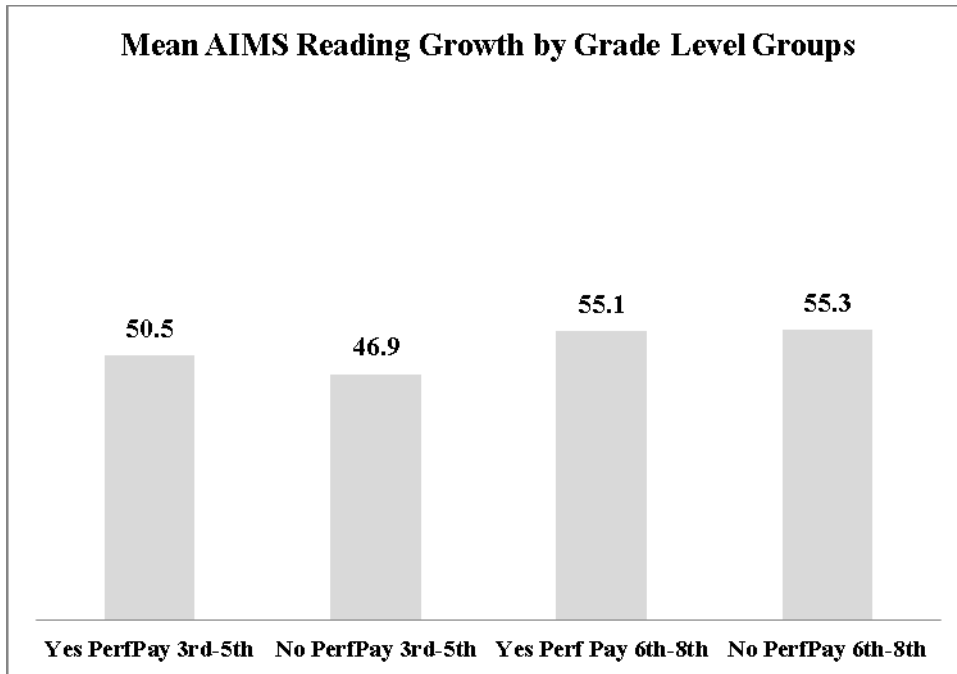|	Count	Percentage of means for AIMS reading growth scores	Standard deviation
3rd-5th Performance Pay Yes	42	50.5	10.6
3rd-5th Performance Pay No	73	46.9	9.6
6th-8th Performance Pay Yes	46	55.1	6.9
6th-8th Performance Pay No	80	55.3	6.8

Figure 6. Percentage of means for AIMS reading growth scores by grade-level groups

Research Question 2

The second research question that guided this research asked, *Is there a statistically significant relationship between teacher evaluation scores and student achievement scores on state achievement tests?* In order to determine if a relationship existed between teacher evaluation scores and student achievement the Pearson Correlation was conducted using SPSS.

Teacher total evaluation scores and the percentage passing AIMS reading were evaluated; the results indicated there was no correlation $r = .124$, $n = 241$, $p = .055$. The second analysis was to see if a relationship existed between total teacher evaluation scores and AIMS reading growth scores. The Pearson results between teacher evaluation scores and AIMS reading growth scores, although slightly higher than the percentage

passing AIMS reading, also demonstrated there was no correlation, $r = .164$, $n = 241$, $p =$ .011.

Further analysis was conducted on the specific rubrics in the teacher evaluation report. These rubrics included lesson content lesson facilitation, student engagement, learning climate, instructional planning, and student academic progress. All rubrics were worth five average points, except instructional planning, which was worth six points, for a total of 31 points possible for the complete teacher evaluation score. Pearson's correlation was completed on all rubrics of the teacher evaluation for both percentages passing AIMS reading and AIMS reading growth scores. The results indicated a very weak correlation between learning climate and AIMS reading growth scores, $r = .207$, $n = 241$, $p = .001$. Student engagement also showed a very weak correlation with AIMS reading growth scores, $r = .189$, $n = 241$, $p = .003$. Lesson content, lesson facilitation, instructional planning, and student academic progress did not have correlations with either AIMS reading passing or AIMS reading growth. Each of the correlation results are detailed in Table 13.

Table 13

Correlation Results for Teacher Evaluation Rubrics and Percentage of Passing AIMS Reading and Growth Scores

	Lesson content	Lesson facili-tation	Student engage-ment	Learning climate	Instruct-ional planning	Student academic progress	Total Eval-uation Points
AIMS percent passing, $n = 241$	$r = .117$ $p = .069$	$r = .122$ $p = .058$	$r = .107$ $p = .099$	$r = .173$ $p = .007$	$r = .055$ $p = .391$	$r = .102$ $p = .116$	$r = .124$ $p = .055$
AIMS growth, $n = 241$	$r = .130$ $p = .045$	$r = .155$ $p = .016$	$r = .189$ $p = .003$	$r = .207$ $p = .001$	$r = .044$ $p = .497$	$r = .180$ $p = .497$	$r = .164$ $p = .011$

Overall, the correlation was higher for AIMS reading growth scores and teacher evaluation rubrics when compared to the percentages for AIMS passing scores, with the exception of the instructional planning rubric. Results indicated a lack of correlation between teacher evaluations and rubric scores and students performance on AIMS reading in both passing percentages and growth calculations.

Summary

Based on the analysis conducted, there was a statistically significant difference on the percentage of students passing the AIMS in reading for teachers who did receive performance pay, especially for teachers in Grades 3 through 5. Teachers' reading growth scores were not statistically significant for teachers awarded performance pay. Scores for teachers sixth through eighth grades for percentages on passing reading and reading growth were not found to be statistically significant for those teachers awarded performance pay. Additional data analysis conducted found no correlation between overall teacher evaluation scores and percentages on passing reading or growth scores. A very weak correlation was found within the specific teacher evaluation rubrics of student engagement and learning climate with AIMS reading growth scores. Correlations were not found between the percentage of students passing on reading AIMS and any portion of the teacher evaluation. Chapter 5 provides a discussion and interpretation of the data including recommendations for future research.

CHAPTER 5

SUMMARY AND CONCLUSIONS

This chapter presents a summary of the important findings and conclusions drawn from the data presented in Chapter 4. Recommendations for further research are also discussed. Changing the existing structure to pay teachers has been the focus for reform efforts over the last 30 years. Teacher salary expenditures represent the majority of budget allocations for school districts, yet simply paying teachers for years of service and education has not yielded higher achievement scores. Performance pay has been viewed as an option to change the structure of teacher compensation with a desired outcome of increased student achievement scores. Arizona, along with 11 other states, was the recipient of Race to the Top funding in 2010 that required school districts to incorporate student achievement data into teacher evaluations to ultimately award highly effective teachers with performance pay. This research sought to determine if a relationship existed in Cardinal Valley School District's current performance pay program between student achievement, teacher evaluation scores, and performance pay.

Summary of the Study

Arizona has changed the components required in teacher evaluations, and districts must now include student achievement outcomes when labeling teachers as effective or ineffective. The inclusion of data and labels of teacher performance are projected to be tied to performance pay by the 2015-2016 school year. School districts all over the state are allowed to develop their own models for the inclusion of data in teacher evaluations.

CVSD is a K through eighth grade urban school district located in Phoenix, Arizona, with 10,118 students enrolled during the 2012-2013 school year. Since 2008,

CVSD has had a performance pay program that has awarded the performance of teachers based on their evaluation scores and their student achievement scores. The district was faced with two distinct populations of teachers: teachers awarded performance pay and teachers not awarded performance pay.

During the 2012-2013 school year, the school district employed a total of 490 teachers. Of the entire population, the sample was refined to 241 teachers in Grades 3 through 8, because these teachers had state reading assessment data that could be directly tied to their classroom instruction. In addition to gathering student-level state reading assessment data for these teachers, their 2012-2013 teacher evaluation scores were collected. The following research questions guided this research:

> *Research Question 1*: Is there a significant difference between the proficiency levels of students on state achievement tests for teachers receiving performance pay when compared to teachers who do not receive performance pay?
>
> *Research Question 2:* Is there a significant relationship between teacher evaluation scores and student achievement scores on state achievement tests?

The district wanted to determine if their existing performance pay program actually produced higher student achievement; and if indeed it did, this framework could be used to meet the new requirements for the state of Arizona. Teachers were grouped into third through fifth grade and sixth through eighth grade levels to account for differences in reading instruction within the different grade levels. In order to examine if teachers who were awarded performance pay actually had higher student achievement

scores when compared to teachers not awarded performance pay, a two-sample

independent *t* test was conducted on the percentage of students passing the AIMS reading

test along with their reading growth scores. In addition, correlations were conducted on

teacher evaluation scores and student achievement scores for both the percentage of the

students who passed the AIMS reading test and for their reading growth scores. For

teachers awarded performance pay compared to teachers not awarded performance pay in

each of the grade-level groups, histograms were developed to demonstrate the mean

percentage for those students passing the AIMS reading test along with their reading

growth scores. Correlations were then competed and presented in tables for each

component of the teacher evaluation rubric and for the percentage of those students

passing the AIMS reading test along with their reading growth scores.

<div align="center">

Summary of Findings and Conclusions

</div>

Performance Pay and Student Achievement

The research conducted was focused on determining if awarding teachers

performance pay actually increased student achievement. The following research

question guided the research: *Is there a significant difference between the proficiency*

levels of students on state achievement tests for teachers receiving performance pay when

compared to teachers who do not receive performance pay?

The first *t* test compared 88 teachers who were awarded performance pay and 153

teachers who were not awarded performance pay by comparing the mean percentage of

those students passing the AIMS reading test. This first test yielded a *p* = .004, indicating

there was a statistically significant difference between the percentage of students passing

the AIMS reading test and the percentage of teachers receiving performance pay.

Teachers awarded performance pay had a mean score of 80.1% passing the AIMS

reading test, and teachers not awarded performance pay had a mean score of 75.3%

passing. When the same *t* test was conducted on the reading growth scores for AIMS, a

p = .200 determined there was not a statistically significant difference in AIMS reading

growth scores for teachers awarded performance pay when compared to teachers not

awarded performance pay. The mean growth scores for teachers awarded performance

pay was 52.9% and teachers not awarded performance pay had a mean growth score of

51.3%. Although it was evident teachers receiving performance pay had a higher

percentage of passing scores, this was not true of their growth scores. Growth scores are

thought to be a better representation of the impact a teacher has during one year of

instruction.

In order to better understand the achievement results, a *t* test was then conducted

on the grade level of the groups to determine if achievement results by grade levels were

different for teachers who had received performance pay when compared to teachers not

receiving performance pay. The results from the third through fifth grade group of

teachers awarded performance pay did show a statistically significant difference

(*p* =.001) in the percentage of students passing the AIMS reading test when compared

with teachers in the third through fifth grade group not awarded performance pay. When

the same analysis was conducted on teachers in Grades 6 through 8, either awarded

performance pay or not, a *p* = .406 demonstrated a lack of statistically significant

difference in the percentages of students passing the AIMS reading test. When the

analysis was conducted to determine if teachers awarded performance pay had higher

student growth scores on the AIMS reading test, the surprising results of a statistically

significant difference could not be found in either of the grade-level groups. Teachers in the third through fifth grade and sixth through eighth grade receiving performance pay did not have statistically significant higher growth scores.

In summary, although an initial analysis demonstrated teachers awarded performance pay had higher percentages as to passing scores, further analysis showed only teachers in Grades 3 through 6 awarded performance pay demonstrated a higher percentage on passing scores. Reading growth scores were not higher for any teachers awarded performance pay when compared to teachers not awarded performance pay.

Data Validity and Student Success

The difference in percentages as to passing AIMS reading tests in Grades 3 through 5 when compared to Grades 6 through 8 could be explained by the amount of time spent during the day on reading instruction. CVSD used a Response to Intervention (RtI) model for reading support. In this RtI model, students in kindergarten through fifth grades meet with teaching staff specifically trained in providing reading interventions for struggling readers, in addition to the 90-minute reading block. Although teachers in Grades 6 through 8 have a 90-minute language arts block (both reading and writing), for all other subjects teachers are expected to incorporate reading into their subject areas. For example, the math and social studies teachers are expected to incorporate reading standards into their instruction in their applicable subject areas. Although the time spent on reading instruction may have had a positive impact on Grades 3 through 5 as to the percentage of those passing reading, this was not seen at all for reading growth scores.

In addition, models such as RtI and cross-departmental integration of standards create data tracking and validity questions. For example, if teachers of a third grade class

who receive performance pay and were provided more reading intervention time, how can it be determined if it was the success on the part of the teacher or on the part of the reading specialist? Additional data must be considered when calculating and linking student success to one teacher. Outside factors influencing student success could be whether students receive tutoring after school from another teacher, at home, or from a privately employed reading tutor.

Another consideration with the difficulty of linking student success to one teacher is the professional learning environment found at CVSD. The school environment is designed to create a system of support for all students to find success. Teachers regularly collaborate on lessons, planning, and data analysis. It is a common practice for many teachers to interchange their students for targeted instruction in peers' grade-level classrooms. For example, one teacher on a fourth-grade team of three teachers may take the group of students receiving the highest scores on the reading test; whereas another teacher will work with the students scoring the lowest on the most recent test. Teachers in CVSD regularly collaborate with their grade-level teams, instructional coaches, and reading interventionists to ensure student success. All of these additional inputs to student success must be considered before linking data to teachers.

The results of the data analysis to determine if teachers awarded performance pay had higher student achievement conducted for CVSD were similar to results found in the Denver ProComp model. The ProComp model introduced value added student growth data to create a better analysis of the impact on one year of teaching. A report on the impact of ProComp by Goldhaber and Walch (2012) found in general increased achievement during ProComp, but inconsistent results were found across grade levels and

subjects. The results found in CVSD were similar to the analysis of ProComp, with results for third through fifth grade teachers higher than teachers in the sixth through eighth grades. Challenges described with ProComp could also be found in CVSD, including how to account for additional variables impacting student achievement, including class-size, student mobility during the school year, and class distribution of special needs students. All these variables must be researched to determine the impact on student achievement before linking performance pay to individual teachers.

Teacher Evaluation Scores and Student Achievement

The second research question sought to determine if a relationship existed between student achievement and teacher evaluation scores. The results and implications are discussed in the remainder of this chapter. The new Arizona framework requires the inclusion of student achievement data along with teacher evaluation scores to create a comprehensive profile used to label teachers. The second research question guided this scenario: *Is there a significant relationship between teacher evaluation scores and student achievement scores on state achievement tests?*

A Pearson correlation was conducted on overall percentages of passing AIMS reading tests, AIMS reading growth scores, and overall teacher evaluation scores. When data analysis was conducted, a result of $r = .124$ indicated no correlation between percentages on AIMS reading tests and overall teacher evaluation scores. When comparing AIMS reading growth scores and teacher evaluation scores, a correlation still did not exist, although the $r = .164$ was slightly higher than the percentage for passing. This was a surprising outcome in the data analysis as it was assumed student achievement would be correlated with teacher evaluation performance data.

57

Further analysis sought to determine if a relationship existed between the individual rubrics on the teacher evaluation and student achievement. The rubrics included lesson content, lesson facilitation, student engagement, learning climate, instructional planning, and student academic progress. After analysis was conducted on all the teacher evaluation rubrics, only student engagement ($r = .189$) and learning climate ($r = .207$) showed weak correlations with AIMS reading growth scores. AIMS reading growth also did not have a correlation with lesson content, lesson facilitation, instructional planning, or student academic progress. No correlation could be found with the percentages of the scores for passing AIMS and any of the teacher evaluation rubrics. The correlation between student engagement and the learning climate could indicate these factors have an impact on student growth. However, the lack of correlation of any teacher evaluation scores on the percentages of passing require additional data and research, especially given the current political trend to incorporate teacher evaluations and student achievement data.

In order to determine why correlations do not exist in the research, the process of teacher evaluation must be considered. Teachers are observed and scored on the various rubrics twice during the school year. The observations are either conducted by site principal or assistant principal, and at some school sites they are completed by both administrators. The week the observation will take place is known to the teacher in advance in order to prepare, but the exact date is not known. The observation must include an uninterrupted lesson of a minimum of 45 minutes; after the observation is complete and notes are analyzed a score is produced.

The teacher observation scores represent only what was observed at that point in time and may not be representative of that teacher's instruction on a daily basis. This creates what could be a skewed data point for the teacher if the lesson was outstanding; but if daily instruction were inadequate, the data could be inflated. If the lesson observed did not go as planned but daily instruction is impeccable, this could be a negatively skewed data point. This is one of the struggles facing educators and the use of teacher evaluation data could represent why a correlation does not exist in the data for CVSD.

Recommendations for Practice

The recommendations for practice are focused on the issue of data validity: how to correctly link student achievement success to teachers for the purpose of performance pay and how to improve the data collected for use in teacher evaluation. In order to correctly link student success for the purposes of performance pay, a comprehensive data tracking system would have to be implemented. Each group involved would have to track the students they serve, including grade levels, interventionists, and individual teachers. This would provide a better framework to link student success to individuals in the school.

The teacher evaluation process needs to be refined to include multiple points of data collected throughout the year. These multiple observations would help create a comprehensive profile of classroom instruction and would be a better representation than one or two visits scored during the year. This revision to the teacher evaluation process would require teacher input and consideration and would require significant time from site administration. Teachers are sometimes nervous about being observed so to increase the observations throughout the year would require training and discussion with all

parties involved. In addition, site administration oftentimes struggle to complete the two required observations during the year, thus adding mandatory additional observations may not be possible. Consideration could be given to employing outside evaluators for additional observations, but then teachers may be concerned with an unknown outside evaluator.

Neither of these solutions will be easy for a district to implement but data validity must be considered, especially when performance pay will be awarded. Awarding teachers for student success that is not actually attributed to their classroom instruction would be a misuse of the award of performance pay for student achievement. In addition, labeling a teacher *effective* on their teacher evaluation based on one classroom observation could also be a misrepresentation of data.

As districts and states move forward with performance pay based on teacher evaluation and student achievement data, it is imperative the data included is valid and reliable. The data should be representative of multiple measures for teachers and students before performance pay should be awarded. In addition to ensuring valid data for teacher evaluation and student achievement data, inclusion on other types of successful performance pay for teachers must be considered.

A successful model for awarding teachers performance pay could include the various types of performance pay, including knowledge-and-skills-based pay coupled with student growth and teacher evaluation scores accumulated over time. Several studies have found a small positive impact for both NBPTS and performance pay for student growth scores (Cavalluzzo, 2004; Clotfelter, Ladd, & Vigdor, 2007; Goldhaber, 2002; Harris & Sass, 2006; Steele, Hamilton, & Stecher, 2010). If teachers were awarded

performance pay for obtaining a NBPTS and additionally awarded performance pay based on student growth over multiple years, this combination could potentially produce higher student achievement over time; longitudinal research would need to be conducted on this topic.

Recommendations for Future Research

Research in the areas of teacher evaluation, performance pay, and linking student achievement is still new and must be ongoing to ensure validity. Future research should include longitudinal studies of teacher evaluation scores and student achievement data. It is possible that looking at multiple years of data would provide a more accurate label for teachers than a single-year snapshot. For example, if a teacher has high student achievement scores and teacher evaluation scores consistently for three to five years, could they be labeled highly effective?

Research needs to include the impact of changes in leadership at schools sites and how that could impact student achievement and teacher evaluation scores. It is possible to have lower or higher scores on a teacher evaluation with a different administrator. Analysis should be conducted to determine if the amount of the performance pay awarded has an impact on student achievement and teacher evaluation scores. For example, if the performance pay is $500 versus $5,000, is there a significant difference in student or teacher evaluation outcomes? Linking student achievement and teacher evaluation scores for the purposes of awarding performance pay must be conducted thoughtfully with integrity and valid data. Given all the money provided to education for the purpose of increasing student achievement, it is clear policy makers need to continue to research best practices in this emerging field. Teachers do not enter the profession with

an expectation of earning a six figure salary; the desire to teach is an intrinsic motivator.

Through continued research on how a teacher earns high achievement scores consistently

over multiple years could provide the answers policy makers seek on how to extrinsically

award teacher performance pay.

REFERENCES

Amator, J., & Kelley, L. (2007). *The effects of the Career Ladder Program on student achievement.* Retrieved from https://www.azed.gov/wp-content/uploads/PDF/CareerLadderReport.pdf

Amrein-Beardsley, A. (2006). Teacher research informing policy: An analysis of research on highly qualified teaching and NCLB. *Essays in Education.* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.9052&rep=rep1&type=pdf

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, *37*(2), 65–75. doi:10.3102/0013189X08316420

Archer, J. (2001). AFT to urge locals to consider new pay strategies. *Education Week*, 20–21. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:AFT+To+Urge+Locals+To+Consider+New+Pay+Strategies#0

Baker, E., Barton, P., & Darling-Hammond, L. (2010). Problems with the use of student test scores to evaluate teachers. *Economic Policy* (Vol. 278, p. 29). Retrieved from http://epi.3cdn.net/b9667271ee6c154195_t9m6iij8k.pdf

Ballou, D., & Podgursky, M. (1997). *Teacher pay and teacher quality*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Boser, U. (2012, March). Race to the top: What have we learned from states so far. *Washington, DC: Center.* Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Race+to+the+Top:+What+Have+We+Learned+from+the+States+So+Far?#0

Bradley, A. (1994). Pioneers in professionalism. *Education Week*, 1–10. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Pioneers+in+Professionalism#0

Brandt, R. M. (1990). *Incentive pay and career ladders for today's teachers.* Albany, NY: State University of New York Press.

Bryant, S. (2002, April 10). Houston School District plans to offer signing bonus again. *Houston Chronicle*.

Cannon, M. D. (2007). *Pay-for-performance: New developments and issues.* Retrieved from

https://my.vanderbilt.edu/performanceincentives/files/2012/10/200705_Cannon_PFP_NewDevel1.pdf (p. 20).

Cavalluzzo, L. (2004, November). *Is National Board certification an effective signal of teacher quality?* CNA Corporation. Retrieved from http://files.eric.ed.gov/fulltext/ED485515.pdf

Chait, R., & Miller, R. (2010). *Treating different teachers differently: How state policy should act on differences in teacher performance to improve teacher effectiveness and equity* (pp. 1–5). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Treating+Different+Teachers+Differently+How+State+Policy+Should+Act+on+Differences+in+Teacher+Performance+to+Improve+Teacher+Effectiveness+and+Equity#0

Clifford, G. J. (1989). Man/woman/teacher: Gender, family, and career in American educational history. In D. Warren (Ed.), *Histories of a Profession at Work* (pp. 293–343). New York: Macmillian.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682. doi:10.1016/j.econedurev.2007.10.002

Conley, S., & Odden, A. (1995). Linking teacher compensation to teacher career development. *Educational Evaluation and Policy Analysis*, *17*(2), 219–237. Retrieved from http://epa.sagepub.com/content/17/2/219.short

Crandall, F., & Wallace, M. (1998). *Work and rewards in the virtual workplace: A new deal for employers and employees*. New York: American Management Assocation, Inc.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012, March). Evaluating teacher evaluation. *Phi Delta,* 93. Retrieved from http://intl.kappanmagazine.org/content/93/6/8.short

Elsbree, W. (1939). *The American teacher: Evolution of a profession in a democracy.* New York: American Book Company.

Ferdinand, P. (1998, July 9). Massachusetts weighs giving signing bonuses to teachers. *Washington Post*, p. A08.

Figlio, D. N. D., & Kenny, L. L. W. (2007, April). Individual teacher incentives and student performance. *Journal of Public Economics*, *91*, 901–914. doi:10.1016/j.jpubeco.2006.10.001

Gewertz, C. (2001, July). Nevada: Pay raises, signing bonuses on tap for state's teachers. *Education Week*, *20*(42), 25.

Glass, G., & Hopkins, K. (1970). *Statistical methods in education and psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Glazerman, S., & Booker, K. (2008). Does the Missouri Teacher Career Ladder Program Raise Student Achievement? Washington, DC: Mathematica Policy Research Inc. Retrieved February 22 2012, from http://www.mathematica-mpr.com/publications/pdfs/education/MCL_achievement.pdf

Goldhaber, D. (2002). The mystery of good teaching. *Education Next*. Retrieved from https://www.stcloudstate.edu/tpi/initiative/documents/preparation/The Mystery of Good Teaching.pdf

Goldhaber, D., & Walch, J. (2012). Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative. *Economics of Education Review*, *31*(6), 1067–1083. doi:10.1016/j.econedurev.2012.06.007

Goldrick, L. (2002). *Improving teacher evaluation to improve teaching quality* (Issue Brief). Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED480159

Gonring, P., Teske, P., & Jupp, B. (2007). *Pay-for-performance teacher compensation: An inside view of Denver's ProComp Plan*. Cambridge, MA: Harvard Education Press. Retrieved from https://tcrecord.org/books/abstract.asp?ContentId=15022

Gratz, D. B. (2009). *The peril and promise of performance pay: Making education compensation work* (Kindle Edition). Rowman & Littelfield Education. Retrieved from http://www.amazon.com/gp/product/B002CQTW90/ref=kinw_myk_ro_title

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*, 1141–1177.

Hanushek, E. A., Fildes, V., & Davies, R. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, *100*(1), 84–117. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Trade-off+between+Child+Quantity+and+Quality#0

Harris, D. N., & Sass, T. (2006). *Value-added models and the measurement of teacher quality.* Unpublished Manuscript. Retrieved from http://myweb.fsu.edu/tsass/Papers/IES Harris Sass EPF Value-added 14.pdf

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7-8), 798–812. doi:10.1016/j.jpubeco.2010.11.009

Heneman, R. L., Ledford, G. E., & Gresham, M. T. (2002). *The changing nature of work and its effects on compensation design and delivery.* In S. Rynes & B. Gerhart (Eds.), *Compensation in organizaions: Current research and practice* (pp. 195–240). San Francisco: Jossey-Bass. Retrieved from http://amr.aom.org/content/26/1/128.short

Humphrey, D., Gallagher, H., & Yee, K. (2012). *Teacher incentive fund: First implementation report, 2006 and 2007 grantees.* Office of Planning.. Retrieved from http://files.eric.ed.gov/fulltext/ED531980.pdf

Johnson, S., & Birkeland, S. (2003). Pursuing a "sense of success": New teachers explain their career decisions. *American Educational Research.* Retrieved from http://aer.sagepub.com/content/40/3/581.short

Johnson, S. M., Kardos, S. M., Kauffman, D., Liu, E., & Donaldson, M. L. (2004, October 29). *The support gap: New teachers' early experiences in high-income and low-income schools.* Education Policy Analysis Archives. Retrieved from http://epaa.asu.edu/ojs/article/view/216

Kelley, C., III, & Heneman, H., III., Milanowski, A. (2000, April). *School-based performance award programs, teacher motivation, and school performance: Findings from a study of three programs* (CPRE Research Report). Consortium for Policy Research in Education. Retrieved from http://files.eric.ed.gov/fulltext/ED477654.pdf

King, R. A., Swanson, A. D., & Sweetland, S. R. (2003). *School finance: Achieving high standards with equity and efficiency*. Needham Heights, MA: Allyn & Bacon.

Koedel, C., & Betts, J. (2010, February). Value added to what? How a ceiling in the testing instrument influences value-added estimation. *Education.* Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/edfp.2009.5.1.5104

Koretz, D. (2002, July). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources.* Retrieved from http://www.jstor.org/stable/3069616

Lawler, E. E. (2000). *Rewarding excellence: Pay strategies for the new economy*. San Francisco: Jossey-Bass.

Liu, E., Johnson, S. M., & Peske, H. G. (2004). New teachers and the Massachusetts signing bonus: The limits of inducements. *Educational Evaluation and Policy Analysis*, *26*(3), 217–236. doi:10.3102/01623737026003217

McGuinn, P. (2012, November). *The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher-evaluation systems.* Center for American Progress. Retrieved from http://files.eric.ed.gov/fulltext/ED539744.pdf

Incentive. (2014). In *Merriam-Webster's online dictionary.* Retrieved from http://www.merriam-webster.com/dictionary/incentive

Mihaly, K., & McCaffrey, D., Staiger, D. O., & Lockwood, J. R. (2013). *A composite estimator of effective teaching* (Technical report). Seattle, WA: Bill &Melinda Gates. Retrieved from http://www.nbexcellence.org/cms_files/resources/

National Board. (n.d.). *National Board for Professional Teaching Standards.* Retrieved January 3, 2014, from http://www.boardcertifiedteachers.org/in-your-state

National Center for Education Statistics. (n.d.). *Schools and Staffing Survey* (SASS). Retrieved from http://nces.ed.gov/surveys/sass/tables/sass1112_2013311_d1n_005.asp

Neil, K. Jr, & Martinez, B. (2014, March 5). Race to the Top finalists are named. *Wall Street Journal*, pp. 8–9.

No Child Left Behind Act, Public Law 1–670 (2001). 107th Congress. Retrieved from http://www.fldoe.org/eias/databaseworkshop/pdf/nclbspar.pdf

Novicoff, W. (2006). Pay-for-performance issues. *Hospitals & Health Networks/AHA*, *80*(5), 8, 10. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3535413&tool=pmcentrez&rendertype=abstract

Odden, A., & Kelley, C. (2002). *Paying teachers for what they know and do* (2nd ed.). Corwin Press.

Podgursky, M., & Springer, M. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management, 26*(4), 909–949. doi:10.1002/pam

Protsik, J. (1996). History of teacher pay and incentive reforms. *Journal of School Leadership*, *6*, 265–289.

Schemo, D. J. (2002, July 6). For Oklahoma's teachers big D is dollars (and Dallas). *New York Times*, p. A7.

Slavin, R. E. (2008). Perspectives on evidence-based research in education—what works? Issues in synthesizing educational program evaluations. *Educational Researcher*, *37*(1), 5–14. doi:10.3102/0013189X08314117

Spencer, D. (2001). Teachers' work in historical and social context. In V. Richardson, L. Corno, M. Dilworth, M. Eisenhart, R. Floden, P. Lather, . . . K. Zeichner (Eds.), *Handbook of research on teaching* (4th ed., p. 1278). Washington DC: American Educational Research Association.

Springer, M. G. (2009). *Performance incentives: Their growing impact on American K-12 education* (p. 336). Washington DC: The Brookings Institution.

Springer, M.G., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, D., Pepper, M., & Stecher, B. (2010a). *Teacher pay for performance: Experimental evidence from the project on incentives in teaching.* Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Springer, M. G., & Gardner, C. D. (2010b, May). Teacher pay for performance context, status, and direction. *Phi Delta Kappan.* Retrieved from http://intl.kappanmagazine.org/content/91/8/8.short

Steele, J., Hamilton, L., & Stecher, B. (2010). *Incorporating student performance measures into teacher evaluation systems* (Technical Report). RAND Corporation. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED514528

U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Race+to+the+Top+Program+Executive+Summary#0

U.S. Department of Education. (2012). *Teacher incentive fund*. http://www2.ed.gov/programs/teacherincentive/2007-awards.html

Vinovskis, M. (1999). *The road to Charlottesville: The 1989 education summit*. Retrieved from http://books.google.com/books?hl=en&lr=&id=DFvw9L3nTKEC&oi=fnd&pg=PA1&dq=The+Road+to+Charlottesville+The+1989+Education+Summit&ots=19BHWQTOMw&sig=P-lazG5Nwf7ElaATAcCF-GOeN5A

Wiley, E., Spindler, E., & Subert, A. (2010). *Denver ProComp: An outcomes evaluation of Denver's alternative teacher compensation system 2010 report.* Boulder, CO: University of Colorado at Boulder. Retrieved from http://static.dpsk12.org/gems/*newprocomp*/ProCompOutcomesEvaluationApril2010final.pdf

Yeh, S. S. (2009). The cost-effectiveness of raising teacher quality. *Educational Research Review*, *4*(3), 220–232. doi:10.1016/j.edurev.2008.06.002

Zingheim, P. K., & Schuster, J. R. (2000). *Pay people right! Breakthrough reward strategies to create great companies*. San Francisco: Jossey-Bass.