

STETSON UNIVERSITY

*Voices of Reform: Educational Research to
Inform and Reform*

Volume 3 • Issue 2 • Article 1



December 2020

Examining Student Achievement, Teacher Effectiveness, and Merit Pay in a Rural Tennessee School District

Mary Boudreaux
Southern Connecticut State University
Jill Faulkner
Chester County Schools

Follow this and additional works at: <http://www.voicesofreform.com>

Recommended Citation

Boudreaux, M. & Faulkner, J. (2020). Examining student achievement, teacher effectiveness, and merit pay in a rural Tennessee school district. *Voices of Reform*, 3(2), 11-39. Retrieved from <https://www.voicesofreform.com/article/18592-examining-student-achievement-teacher-effectiveness-and-merit-pay-in-a-rural-tennessee-school-district> doi: 10.32623/3.10003

<http://dx.doi.org/10.32623/3.10003>

Revisions

Submission date: March 22nd, 2020
1st Revision: July 13th, 2020
Acceptance: July 20th, 2020
Publication date: December 29th, 2020

Examining Student Achievement, Teacher Effectiveness, and Merit Pay in a Rural Tennessee School District

Mary Boudreaux¹
Jill Faulkner²

¹Department of Educational Leadership and Policy Studies
Southern Connecticut State University, United States
boudreauxm1@southernct.edu

²Chester County School District Office
Chester County Schools, Henderson, Tennessee, United States
jill.faulkner@chestercountyschools.org

Abstract

With more rigorous standards and testing at the forefront of educational reform across the nation, the rural school district in this study developed a strategic compensation plan with bonus pay based on student test scores as a teacher success incentive. A causal-comparative study was conducted to examine the effect of teacher merit pay levels on teacher effectiveness and student achievement within a rural school district. The study also considered if there is a difference among the teacher effectiveness levels and student achievement scores and certain moderating variables including content area, years of experience, gender, and education level. One-way ANOVA analysis determined student achievement scores were significantly lower when the teacher did not qualify for a bonus than at any and every other bonus level. Linear regression analyses found significant moderation effects for years of experience, education level, and content area, but not for teacher gender.

Keywords

merit pay, teacher compensation, student achievement, educational policy

Introduction

Accountability mandates at the national level prodded states to seek ways to improve teacher effectiveness and student achievement. The latest PISA results still show the United States in the middle of the rankings for student achievement among the countries members of the Organization

of Economic and Co-operation and Development (OECD) (PISA, 2015). As a result, policymakers are anxious to discover ways to improve student achievement so that Americans can compete in a global economy. The federal government has exercised control over schools through the regulation of funding embedded in programs such as No Child Left Behind (NCLB), Race to the Top (RTTT), and the Every Student Succeeds Act (Spring, 2016) in an attempt to spur schools to increase student achievement. While education is a function of each state, these accountability policies enacted by the federal government imposed many requirements on the states. These requirements triggered reform at the state level in teacher evaluation models, curriculum standards, teacher pay structures, student testing, and educator preparation programs. According to Dee and Wyckoff (2017), two out of three states revamped their teacher evaluation policies between 2009 and 2015 due to federal incentives through Race to the Top, the Teacher Incentive Fund, and No Child Left Behind. Tennessee ranked among the lowest five states for student achievement on the math portion of the National Assessment of Education Progress (NAEP) test in 2009 for both fourth and eighth grades and 39th and 34th, respectively, in fourth and eighth grade Reading (National Assessment of Educational Progress, 2009). The school district in this study scored below the state average on the 2011 Tennessee Comprehensive Assessment Program (TCAP) end of course (EOC) tests in Algebra I, Algebra II, and English I (Tennessee Department of Education, TDOE, 2011). English II scored only marginally higher than the state average (TDOE, 2011). As a result, one way the district in this study sought to improve student test scores was through improving teacher effectiveness through a federally funded competitive grant, Teacher Incentive Fund.

Additionally, policymakers examined using private sector ideals in the management of public education going as far back as the mid-1990s (Mintrop, Ordenes, Coghlan, Pryor, & Madero, 2018). This led to some states implementing teacher merit pay based on including student achievement in teacher evaluation. Since 2010, thirty-six states and Washington, D.C. applied for and received Teacher Incentive Fund grants to improve educator effectiveness through student achievement which affected 2000 schools (Office of Innovation and Improvement, n.d.). Although researchers studied teacher merit pay and student achievement, there is little agreement on the impact of teacher merit pay on student achievement (Ballou, Springer, McCaffrey, Lockwood, Stecher, Hamilton, & Pepper, 2012; Kaimal & Jordan, 2016; Shifrer, Turley, & Heard, 2017). Hendricks (2014) cited a lack of strong evidence that merit pay influenced student achievement. He stated that such studies should control for changes in the district characteristics that could affect student achievement (Hendricks, 2014). In contrast, a short-term effect of teacher bonus programs included an increase in teacher effectiveness and thereby also, student achievement, in an effort by teachers to increase their pay and decrease the public threats of accountability (Belfield & Heywood, 2008; Lauen, 2013). Long-term effects of merit pay plans are they could change the composition of the teaching workforce (Bowen & Mills, 2017; Lauen, 2013; Podgursky & Springer, 2007).

Several studies did not show a significant impact on student achievement including the Chicago Teacher Advancement Program (TAP), Texas Educator Excellence Grant (TEEG), and Minnesota's Q-Comp (Choi, 2015). Denver's Pro-Comp program showed a positive effect in middle school math, but the effect in middle school reading and high school math and reading was insignificant (Goldhaber & Walch, 2012). Springer, Lewis, Podgursky, et al. (2009) found a

negative effect on student achievement in the Governor’s Educator Excellence Grant program in Texas.

In a recent study of the Teacher Incentive Fund, the results indicated that the merit pay program has some effect some effect on improving student achievement and increasing effective teachers, but it is not a persuasive one (Iasevoli, 2017; Pham, Nguyen, Springer, & Vanderbilt University, 2017). Therefore, there is no conclusive evidence at the national level on merit pay as it relates to teacher effectiveness and student achievement.

Merit Pay and Accountability

Hopmann (2011) declared over ten years ago that we are living in an *age of accountability*. Accountability is defined as the process of identifying who is responsible (schools, teachers, and students) for what (learning) (Deming & Figlio, 2016; Editorial Projects in Education Research Center, 2004; Lavigne, 2014; Smith, 2016). Current research on accountability takes a rationalist approach (Alexander, Jang, & Kankane, 2017). This approach blames the “wicked problems” of accountability on a lack of incentives or system supports (Holloway, Sørensen, & Verger, 2017; Rittel & Webber, 1973). Education predominantly employs managerial accountability based on incentives and sanctions (Dubnick, 2006). This particular order of accountability heavily relies on standardized test results (Holloway et al., 2017).

In 1983, *A Nation at Risk* foreshadowed the accountability movement in the United States (Deming & Figlio, 2016). States began ranking K-12 schools using student performance in response to this report. Deming and Figlio (2016) further stated that this movement peaked with the passage of the No Child Left Behind Act (NCLB) of 2001. NCLB required all K-12 schools to test students in core subjects and tie teacher evaluations to those scores (Deming & Figlio, 2016; Editorial Projects in Education Research Center, 2004; Holloway et al., 2017; Hopmann, 2008; Jennings & Lauen, 2016). In December 2015, the *Every Student Succeeds Act* (ESSA) replaced NCLB thereby returning some of the implementation power of accountability to the states and reduced testing (Deming & Figlio, 2016; Jennings & Lauen, 2016). Several factors contribute to accountability frameworks include student achievement, teacher content and pedagogical knowledge, and teacher performance.

Theoretical Foundation

Expectancy Theory

Expectancy theory has been widely used as a theoretical framework for merit pay (Kellough & Lu, 1993). This theory purports that motivation is the function of 3 interrelated factors of valence, expectancy, and instrumentality (Vroom, 1964). Valence is the attractiveness of the reward to the individual (Northouse, 2016; Rice, Malen, Jackson, & Hoyer, 2015; Vroom, 1964). Expectancy is the individual’s belief that their work will lead to the desired outcome (Northouse, 2016; Rice et al., 2015; Vroom, 1964). Finally, instrumentality is the individual’s belief they are capable of performing the work (Northouse, 2016; Rice et al., 2015; Vroom, 1964).

In regards to valence, teachers participating in merit pay programs in Texas and Washington, D.C. were more likely to leave a school if they received no or a relatively small bonus than those that

received a large bonus (Dee & Wyckoff, 2015; Springer, Lewis, Podgursky, et al., 2009a). A report by the United States Department of Education (2012) on the first implementation of the Teacher Incentive Fund found that awards less than two percent of an educator's base salary are not sufficient to motivate teachers, but those above four percent may induce the desired behavior. In regards to expectancy and instrumentality, if the educator perceives the reward as attainable and linked to their effort, motivation increases (Kelley, Heneman, & Milanowski, 2002).

Goal-Setting Theory

Goal-setting theory (Locke & Latham, 2006) sets forth that specific, high goals lead to a higher level of task performance than do easy or vague goals. The expectations of this theory are that as long as person is committed to the goals, has the ability to attain it, and does not have conflicting goals, there is a positive linear relationship between the difficulty of the goal and the performance of a task (Locke & Latham, 2006). Goal-setting theory acknowledges program fairness as well as the attainment of goals as important mediating factors (Rice, Malen, Jackson, & Hoyer, 2017). Studies using expectancy and goal-setting theories suggest that awards must have a high valence; awards must address expectancy and instrumentality; and goals, measures, and awards must be perceived as fair (Springer, Lewis, & Podgursky, et al., 2009b; U.S. Department of Education, 2012; Yuan, Le, & McCaffrey, et al., 2013). This study uses expectancy theory and goal-setting theory as the theoretical framework because of the design of the local compensation plan.

Literature Review

Student Achievement

Since before the launch of Sputnik, Americans have compared themselves to other countries. With the issuance of the report *A Nation at Risk*, leaders sought to find a way to improve education, especially in math and science. American student scores on international tests such as PISA and the Trends in International Math and Science Studies (TIMSS) have consistently been lower than other developed countries. While students in the United States have improved in math and science on the TIMSS, they still rank outside of the top quarter (Provasnik et al., 2016). Additionally, per pupil expenditures in the United States has increased from \$5984 in 1970 to \$13, 373 (2015-16 dollars) in 2013 (Deming & Figlio, 2016; National Center for Education Statistics, 2003). As mentioned above, an increase in student achievement scores has been much slower. Therefore, current federal policy made school accountability based on student test scores a priority through the use of ratings, rankings and school report cards, as well as through rewards and sanctions (Deming & Figlio, 2016; Hanushek & Raymond, 2005).

Despite this focus on student achievement, Hanushek and Raymond (2005) found there is lack of consensus on any certain measure of a school's characteristics identified as relevant in determining student performance. However, student achievement increased in response to accountability pressures in Florida's schools partly due to changes in school policy, such as focusing on the lower achieving students, increasing instructional time, and increasing teacher resources (Dee & Jacob, 2011; Dee, Jacob, & Schwartz, 2013; Rouse, Hannaway, Goldhaber, & Figlio, 2007). Deming and Figlio (2016) report a typical finding in the research that accountability increases math test scores in the lowest performing schools by about one-tenth of a standard deviation when compared to the

highest performing schools. Most of the interest in accountability is focused on the lower achieving students (Deming & Figlio, 2016; Hanushek & Raymond, 2005). According to Hanushek and Raymond (2005), accountability seeks to close achievement gaps through the disaggregation of ethnic groups. Consequently, Hispanic students gain the most from accountability while black students gain the least (Hanushek & Raymond, 2005). Jennings and Lauen (2016) point out that accountability has contributed to an increase in the black-white achievement gap. However, Dee and Jacob (2011) found that NCLB had a larger impact on black students and Hispanics in fourth grade math than on white students.

Some concerns related to the use of student test scores in accountability policies include cheating, teaching to the test, and an undesirable narrowing of the curriculum (Deming & Figlio, 2016; Hanushek & Raymond, 2005; Jennings & Lauen, 2016). Other consequences of the emphasis on student test scores are focusing on memorization instead of learning, focusing on the borderline students, spending instructional time on test preparation, and manipulating the testing pool by temporarily or permanently removing students from the school (Smith, 2016). Therefore, accountability policies utilizing student outcomes in the form of achievement test scores can have unintended effects other than as motivation to improve teacher effectiveness.

Tennessee's assessment program is called Tennessee Comprehensive Assessment Program (TCAP). Assessments in grades K-2 are optional according to the Tennessee Department of Education website. Grades 3-8 take achievement tests in math, ELA, and science and grades 9-12 take end of course (EOC) tests in English I, II, III, Algebra I, II, Biology, and Chemistry. High School Policy 2.103 requires students to take end of course exams in the semester they complete the relevant course. Students are not required to pass any one exam, but must achieve a passing score for the course (Tennessee State Board of Education, 2018a). If results are received within five days of the end of the course, the score counts for 25% of the second semester final grade (Tennessee State Board of Education, 2018a). Additionally, students with 504 plans or IEPs must be provided the necessary accommodations (Tennessee State Board of Education, 2018a).

Teacher Content and Pedagogical Knowledge and Student Achievement

In his synthesis of over 800 meta-analyses on achievement, Hattie (2009) found that what matters most in whether students learn is teacher expertise. The literature on content knowledge of teachers stresses the importance of specific content knowledge (CK) as well as pedagogical content knowledge (PCK) of teachers. Shulman (1986) defined subject matter knowledge as the conceptual understanding of a subject. Teacher subject knowledge exerts a statistically significant impact on student achievement (Metzler & Woessmann, 2010). Metzler and Woessmann (2010) found that a one standard deviation increase in teacher subject specific academic test scores raises student test scores by about 9% of a standard deviation in math. The effect in reading scores are significantly less and mostly indistinguishable from zero (Metzler & Woessmann, 2010).

Pedagogical content knowledge is defined by Shulman (1986) as the knowledge needed to make subject matter accessible to students. PCK encompasses specific content and instructional strategies as well as an understanding of student misconceptions and learning difficulties (Fernandez, 2014). "Recent studies have provided strong, representative evidence that teachers'

subject-matter knowledge affects their instructional practice and their students' achievement gains" (Kleickmann et al., 2013, p. 91). Both PCK and CK were found to have an effect on student learning, however, CK had lower predictive power for student growth than did PCK (Baumert et al., 2010). Keller, Neumann, & Fischer (2017) assume that teachers with a high PCK can develop a learning environment that not only challenges but also supports students' learning by anticipating student difficulties and responding adaptively to student problems. In her 1999 comparative study between the United States and China, Liping Ma found the level of mathematical content knowledge influences the pedagogical skill of the teacher to analyze student errors or develop conceptual understanding in math (Depaepe, Verschaffel, & Kelchtermans, 2013; Loewenberg Ball, Thames, & Phelps, 2008; Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013; Tchoahanov, 2011; Toh, 2017).

Teacher Performance

The importance of teachers in the education of students has long been emphasized (Aaronson, Barrow, & Sander, 2002; Alexander et al., 2017; Goldhaber, 2002; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004; Wright, Horn, & Sanders, 1997). In fact, teacher effectiveness is the single most important in-school factor in determining student achievement (Winters, 2011). As a result, accountability policies often use student achievement as a measure of teacher effectiveness. However, Harris and Herrington (2006) contend that these accountability policies rarely closed achievement gaps because few increased student access to resources or content. In contrast, Dee and Wycoff (2015) found that large financial rewards based on several indicators of teacher effectiveness coupled with the threat of dismissal resulted in improvements in teacher effectiveness. Standards-based performance teacher evaluations are more accepted when teachers value the evaluation criteria as related to their own internal standards of quality, when they believe the system is fair, when they believe they have the capability to reach satisfactory performance, when they consider the evaluator credible and trustworthy, and when they receive specific feedback on how to improve performance (Archer, Kerr, & Pianta, 2014; Hatry, Greiner, & Ashford, 1994; Kimball, 2002; Milanowski & Heneman, 2001; Odden, Kelley, Heneman, et al., 2001; Range, Young, & Hvidston, 2013; Stiggins & Duke, 1988).

Despite this lack of agreement on using student test scores in teacher evaluation systems, many states now include student assessment measures in teacher evaluations. In 2017, 39 states required student growth measures to be included in teacher evaluations which is down from 43 states in 2015 (National Center for Teacher Quality, 2017). Indiana is currently the only state to require teachers be rated as effective on student growth measures to receive an overall effective rating (NCTQ, 2017). Consequently, only Hawaii, Nevada, and New York stipulate that teachers that are rated ineffective on student growth cannot receive an overall effective rating (National Center for Teacher Quality's, 2017). In Tennessee, student growth measures can account for as little as 15% to as much as 35% of a teacher's overall evaluation score (Tennessee Department of Education, 2013).

Darling-Hammond (2015) identified several ideal conditions that must exist to accurately pinpoint an individual teacher's contribution to student learning: assessments that measure essential learning and individual student achievement, random assignment of students to teachers, and only

one teacher responsible for student learning during the time period assessed. However, it is rare that all these conditions are met simultaneously. Most standardized tests assess grade level standards and do not accurately measure students that are above or below grade level (Darling-Hammond, 2015). Teachers only account for less than 10% of student learning with school factors such as class size, curriculum, etc.; prior teachers and schooling; peer influences; summer gains and losses in learning; home factors; and individual student needs accounting for the majority of the influence on student learning (American Statistical Association, 2014; Darling-Hammond, 2015).

In a study of the District of Columbia Public School's (DCPS) high-stakes evaluation system, IMPACT, teacher effectiveness was increased through teacher turnover of the lowest performing teachers while having a positive impact on student achievement (Dee & Wyckoff, 2017). DCPS contributes this success to the "high-powered individually targeted incentives linked to performance" (Dee & Wyckoff, 2017, p. 66). In a review of states including student achievement in teacher evaluations over a six-year period from 2007-2013, there was a moderate association with an increase in reading achievement but no significant association with math achievement (Alexander et al., 2017).

To improve student achievement through teacher effectiveness, many states have used federal money to implement teacher merit pay plans. It is the goal of this research study to examine such approach on a local level.

Method

This quantitative study was *ex post facto*, causal-comparative in nature with the purpose of analyzing the effects of a strategic compensation plan on teacher effectiveness levels and student achievement scores. Creswell (2017) points out that causal-comparative research compares two or more groups (levels of teacher of merit pay) related to a situation that has already happened (i.e. *ex post facto*). McMillan (2016) defines causal-comparative design as having an intervention without direct control. Ex post facto research investigates an intervention that has already occurred (McMillan, 2016). Since the merit pay has already occurred, this study fits this research design. The problem was the expectation that student test scores and teacher effectiveness levels would drop with the implementation of more rigorous standards and state assessments brought about by the *First to the Top* Act. Therefore, the district designed a differentiated compensation program to incentivize teachers to improve student achievement scores based on the tenets of expectancy and goal-setting theory. By using a causal-comparative research design, the effect of the merit pay bonus level can be compared to student achievement scores and teacher effectiveness levels. The following research questions were addressed:

1. Is there a statistically significant relationship between teacher merit pay level and teacher effectiveness levels (significantly below expectations, below expectations, at expectations, above expectations, and significantly above expectations) in a school district?
2. Is there a statistically significant relationship between teacher merit pay level and student achievement scores in a school district?

3. Do teacher demographic characteristics (content area, years of experience, education level, and gender) moderate the association between merit pay level, teacher effectiveness level, and student achievement scores in a school district?

Setting and Sample

The setting is a rural school district in Southwest Tennessee. The population is all the teachers and students in the district. Using purposeful sampling, students enrolled in English I, English II, Algebra I, or Algebra II and teachers of these subjects in the years from 2011-2015 at the only high school in the district will be used for the sample (n=3340). Purposeful sampling will be used since the study is focused on one district and the relationship between its differentiated pay plan and student achievement and teacher effectiveness. Students with a valid achievement test score and teachers that received a level of effectiveness during the years studied were eligible for the study.

The district in the study employs 166 classroom teachers and has a student enrollment of 2,819 (TDOE, 2017). The breakdown by gender is almost equal with 50.8% male students and 49.2% female students (TDOE, 2017). The percentage of students considered economically disadvantaged is 31.9%, students with disabilities is 11%, and the English learners' percentage is 0.2% (TDOE, 2017).

The high school involved in the study has a student enrollment of 903 students and 48 classroom teachers (TDOE, 2017). The percentage of economically disadvantaged students is 26.7%, students with disabilities is 9%, and English learners is 0.1% at the school level. The academic achievement indicator on the State Report Card measures the percentage of students performing on grade level on the state assessments and the improvement in percentage from one year to the next. The high school in this study scored 0.6 out of a possible 4 points on the academic achievement indicator (TDOE, 2017). The success rate for the school is 30.3% which is below both the district (39.1%) and state (41%) averages (TDOE, 2017).

Using purposeful sampling, students enrolled in English I, English II, Algebra I, or Algebra II and teachers of these subjects in the years from 2011-2015 at the only high school in the district yielded a total dataset size of N=3340. The district in the study employs 166 classroom teachers and has a student enrollment of 2,819 (Tennessee Department of Education (TDOE), 2017). The breakdown by gender is almost equal with 50.8% male students and 49.2% female students (TDOE, 2017). The percentage of students considered economically disadvantaged is 31.9%, the percentage of students with disabilities is 11%, and the English learners' percentage is 0.2% (TDOE, 2017). Table 1 summarizes demographic, effectiveness, and payment statistics for teachers in the study dataset, separated by year. Teacher gender shows the number of student scores of male teachers and number of student scores by female teacher, respectively. This table highlights that there were no teachers in this sample who scores "significantly below expectations," none that had 16-20 years of experience or more than 26 years of experience, and none with degrees higher than a Master's degree. It also shows that bonus levels were increased after the first year of implementation.

Table 1: *Teacher Demographic, Effectiveness, and Payment Statistics*

	All Data		2014		2015		2016		2017	
	(N = 3245)		(N = 789)		(N = 768)		(N = 824)		(N = 864)	
	N	%	N	%	N	%	N	%	N	%
Gender^a										
Male	878	27.1	274	34.7	320	41.7	133	16.1	151	17.5
Female	2367	72.9	515	65.3	448	58.3	691	83.9	713	82.5
TLOE^b										
Below expectations	342	10.5	223	28.3	119	15.5	0	0.0	0	0.0
At expectations	964	29.7	89	11.3	236	30.7	253	30.7	386	44.7
Above expectations	522	16.1	106	13.4	0	0.0	243	29.5	173	20.0
Significantly above expectations	1417	43.7	371	47.0	413	53.8	328	39.8	305	35.3
Amount of Bonus^c										
None	738	22.7	155	19.6	0	0.0	224	27.2	359	41.6
\$237, \$400	1072	33.0	297	37.6	185	24.1	207	25.1	383	44.3
\$356, \$650	267	8.2	0	0.0	119	15.5	148	18.0	0	0.0
\$533, \$900	782	24.1	122	15.5	293	38.2	148	18.0	122	14.1
Years of Experience^d										
0-5 years	1488	45.9	418	53.0	356	46.4	387	47.0	327	37.8
6-10 years	1048	32.3	330	41.8	292	38.0	187	22.7	239	27.7
11-15 years	387	11.9	0	0	0	0.0	180	21.8	207	24.0
21-25 years	322	9.9	41	5.2	120	15.6	70	8.5	91	10.5
Level of Education^e										
Bachelor's degree	1790	55.2	358	45.4	337	43.9	561	68.1	534	61.8
Master's degree	1455	44.8	431	54.6	431	56.1	263	31.9	330	38.2
Content Area										
Algebra 1	680	21.0	180	22.8	120	15.6	188	22.8	192	22.2

Algebra 2	791	24.4	185	23.4	214	27.9	179	21.7	213	24.7
English 1	673	20.7	212	26.9	223	29.0	235	28.5	226	26.2
English 2	667	20.6	212	26.9	211	27.5	222	26.9	233	27.0
Student Perf. Level										
Below basic	763	23.5	96	12.2	57	7.4	374	45.4	236	27.3
Basic/Approaching	1032	31.8	195	24.7	162	21.1	259	31.4	416	48.1
Proficient/On track	1158	35.7	386	48.9	415	54.0	160	19.4	197	22.8
Advanced/Mastered	292	9.0	112	14.2	134	17.4	31	3.8	15	1.7

Note: ^a The gender represents the number of student scores of male teachers and number of student scores by female teachers. ^bNo teachers scored “significantly below expectations”. ^cThe bonus levels were increased after the first year of implementation. ^dThere were no teachers that had 16-20 years or over 26 years of experience. ^eThere were no teachers with above a Master’s degree.

Table 2: *Descriptive Student Achievement Score Summaries Separated by Year*

	N	Min	Max	Mean	Std. Dev.
Total	3245	0.00	553.00	72.0977	22.36170
2014	789	53.00	100.00	84.98	8.72
2015	768	0.00	99.00	86.34	8.40
2016	824	0.00	97.00	71.29	13.45
2017	864	4.00	96.00	47.53	16.87

Table 2 summarizes descriptive score variables separated by year. Across all years examined in this study, student achievement scores ranged between 0 and 100, with an average score of 71.87 and a standard deviation of 20.15. The standard deviation represents the spread of the sample from the mean and indicates that approximately two-thirds of the scores in this sample were within 20.15 points below and above the mean. Although the maximum scores stayed relatively constant across all years of data included in this study, the minimum score in 2014 was 53, while it was 0 or close to 0 for the other years. The means and standard deviations stayed relatively constant across all years of the study.

Instrumentation

The instrument used for student achievement measures was the Tennessee Comprehensive Assessment Program (TCAP) End-Of Course state achievement tests. The EOCs are a criterion referenced test and the validity and reliability are verified by the Tennessee Department of Education. The results of the EOCs are used for diagnostic purposes to assess student achievement and to provide data in support of improved student instruction. The data analyzed included four years of testing after the implementation of a merit pay program.

Results of Statistical Analysis

Data were entered and stored using the Statistical Programming for the Social Sciences (SPSS) software package version 25. Demographic and descriptive tables were created in SPSS as well. Final statistical analyses and accompanying assumption validation analyses were performed in SPSS version 25.

Research Question 1. To address the first research question, examining the relationship between teacher effectiveness levels (significantly below expectations, below expectations, at expectations, above expectations, and significantly above expectations) and teacher merit pay level in a school district, a chi-square analysis was conducted. This analysis allows for an examination of whether a larger proportion of teachers who receive high merit pay are in the higher effectiveness groups compared to the proportion of teachers who receive little or no merit pay. As RQ1 addresses two categorical groups that are independent from each other and have two or more levels, all assumptions for chi-squared analysis were met.

Table 3: *Cross Tabulation Table for Level of Effectiveness and Level of Bonus Pay*

Teacher Level of Effectiveness		Bonus Level ^a				Total
		0	1	2	3	
Below expectations	Count	155.00	0.00	119.00	0.00	274.0
	Expected count	70.7	102.7	25.6	74.9	274.0
At expectations	Count	354.00	439.00	0.00	0.00	793.0
	Expected count	204.7	297.3	74.1	216.9	793.0
Above expectations	Count	146.00	122.00	148.00	0.00	416.0
	Expected count	107.4	156.0	38.8	113.8	416.0
Significantly above expectations	Count	83.00	511.00	0.00	782.00	1376.0
	Expected count	355.2	515.9	128.5	376.4	1376.0

Note: ^a 2011: 0=no bonus, 1=\$237, 2=\$356, 3=\$533; 2012 and beyond: 0=no bonus, 1=\$400, 2=\$650, 3=\$900

Table 4: *Chi-Square Summary*

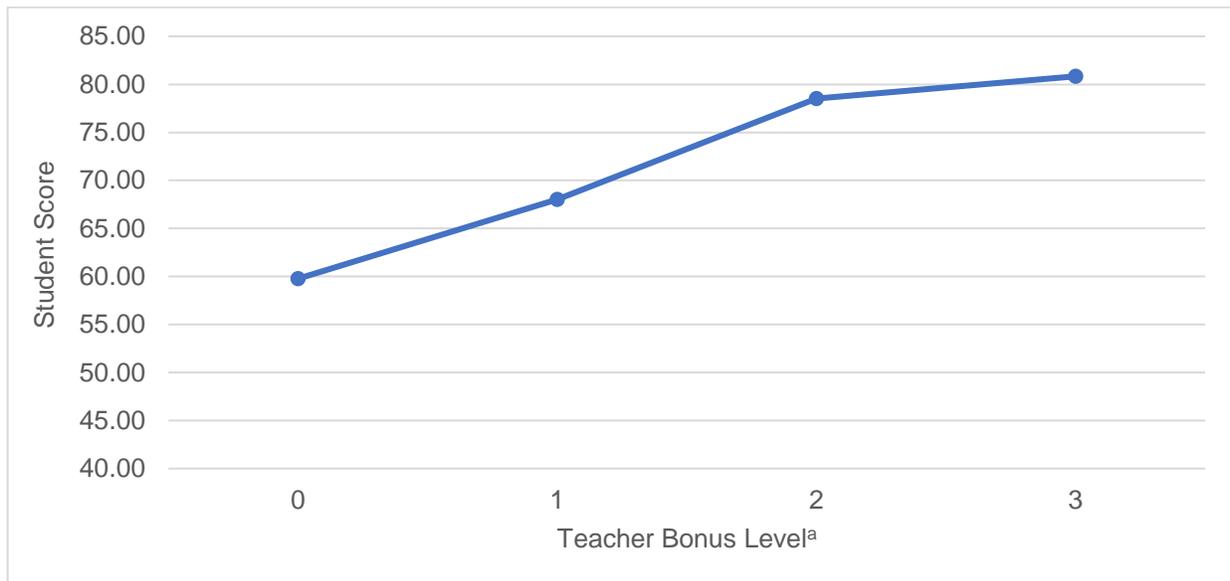
	Value	Df	P
Pearson Chi-Square	2302.484	9	<.001
Likelihood Ratio	2657.144	9	<.001
Linear-by-Linear Association	844.979	1	<.001

As no teacher scored ‘significantly below expectations’, only teacher effectiveness levels of “below expectations,” “at expectations,” “above expectations,” and “significantly above expectations” are displayed in Table 3. Chi-Square test revealed a highly significant difference between expected and observed counts comparing the two categorical groups ($\chi^2 = 2302.48$ $p < .001$). The table summary indicates that considerably more teachers at the highest effectiveness level (significantly above expectations) also belong to higher bonus levels than what would be expected in a normal distribution. This indicates that higher teacher effectiveness may have a relationship with higher bonuses.

Research Question 2. To address the second research question, examining the relationship between teacher merit pay levels and student achievement scores, a one-way ANOVA was conducted. Level of bonus pay served as the independent variable and student achievement scores as measured by the TCAP End-Of Course state achievement tests (percent of questions that each student answered correctly) served as the continuous dependent variable.

Average student achievement scores separated by bonus levels are displayed in Figure 1. Table 5 summarizes the results of the one-way ANOVA and accompanying Levene’s heterogeneity test (Table 6). Table 7 depicts the summary table for post-hoc pairwise comparison using Tukey’s Honest Significant Difference test.

Figure 1: Average Student Achievement Scores Separated by Bonus Levels



Note: ^a 2011: 0=no bonus, 1=\$237, 2=\$356, 3=\$533; 2012 and beyond: 0=no bonus, 1=\$400, 2=\$650, 3=\$900.

Table 5: ANOVA Summary of Student Achievement Score by Bonus Level

	Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>P</i>
Between Groups	192255.83	3	64085.28	137.24	<.001
Within Groups	1333131.55	2855	466.95		
Total	1525387.38	2858			

Table 6: Levene's Test of Homogeneity of Variance, Welch, and Brown-Forsythe Tests

	Statistic	<i>df</i> ₁	<i>df</i> ₂	<i>P</i>
Levene's Test				
Based on Mean	85.15	3	2855.00	<.001
Based on Median	68.32	3	2855.00	<.001
Based on Median and with adjusted <i>df</i>	68.32	3	1801.62	<.001
Based on trimmed mean	86.08	3	2855.00	<.001
Welch Test	226.75	3	1113.047	.000
Brown-Forsythe Test	178.53	3	2498.751	.000

Table 7: Pair-Wise Comparison of Achievement Scores by Bonus Levels

Bonus	Mean Diff	S.E.	<i>P</i>
0 vs. 1	-8.26	1.034	< .001
0 vs. 2	-18.75	1.543	< .001
0 vs. 3	-21.07	1.109	< .001
1 vs. 2	-10.49	1.478	< .001
1 vs. 3	-12.81	1.016	< .001
2 vs. 3	-2.33	1.532	.43

Levene's assumption test indicated a violation of dataset heterogeneity ($p < 0.05$). Therefore, one-way ANOVA was also calculated with Brown-Forsythe and Welch adjustments to determine if controlling for non-homogeneity affected significance levels. All three ANOVAs indicated highly significant differences among bonus levels ($p < 0.001$, Table 6), meaning we can reasonably assume true differences in spite of initial assumption violation. Post-hoc pair-wise comparison indicated that student achievement scores when teachers had no bonus were significantly less than any and every other bonus level (Figure 1 and Table 7).

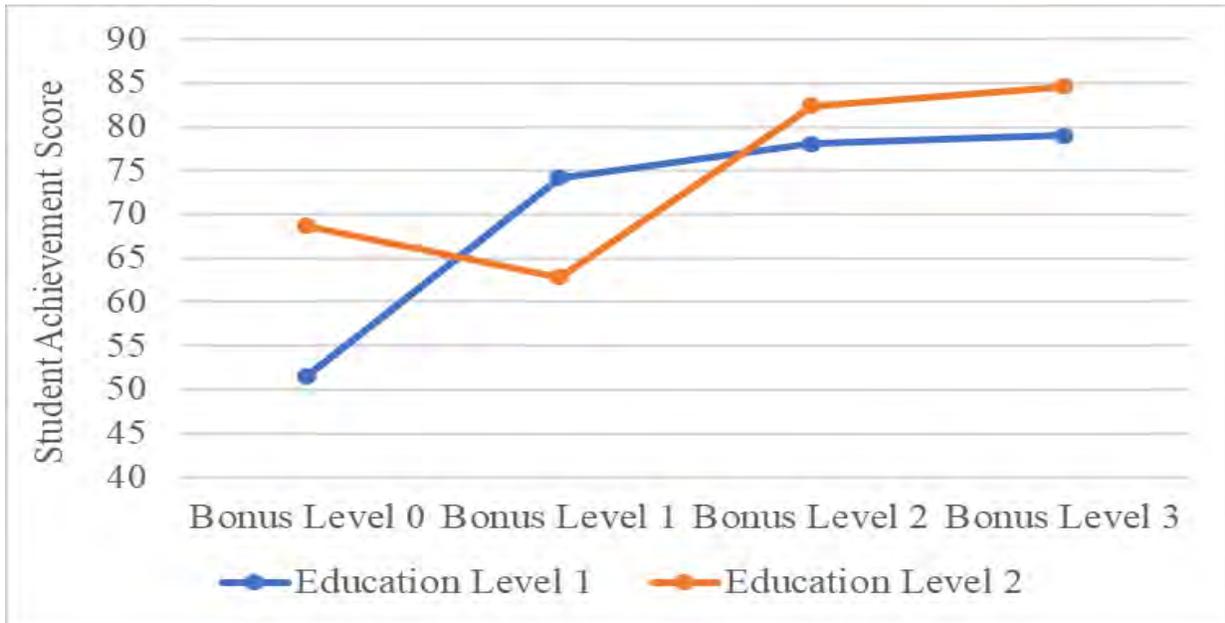
Research Question 3. To address the third research question, a linear regression analysis with moderation was conducted to examine whether the association between bonus level and student achievement scores are moderated by teacher demographics. To create the moderation variables, teacher bonus level was multiplied by each teacher demographic characteristic to create four separate interaction effects (i.e., bonus level*education level, bonus level*gender, bonus level*content level, bonus level*years of experience). Separate linear regression analyses were conducted to examine each potential moderator. In each analysis, teacher bonus level, the teacher demographic characteristic, and the associated moderation variable were entered as independent variables, and student achievement score was entered as the dependent variable. Results of these four separate analyses are shown in Table 8.

Table 8: *The Moderating Role of Teacher Demographic Characteristics*

Model	<i>B</i>	SE	β	<i>t</i>	<i>p</i>	<i>F</i>	Overall <i>p</i>
Bonus Level	5.94	1.50	0.29	3.97	< .001	139.40	< .001
Gender	-4.72	1.34	-0.09	-3.53	< .001		
Interaction	0.82	0.83	0.08	0.99	.32		
Bonus Level	11.528	0.83	0.57	13.98	< .001	184.21	< .001
Years of Experience	-1.157	0.57	-0.06	-2.03	.04		
Interaction	-1.764	0.35	-0.26	-5.08	< .001		
Bonus Level	9.767	1.09	0.48	8.99	< .001	136.57	< .001
Level of Education	3.874	1.28	0.08	3.04	.00		
Interaction	-1.855	0.74	-0.14	-2.50	.01		
Bonus Level	12.026	1.06	0.59	11.40	< .001	189.52	< .001
Content Area	7.020	0.63	0.33	11.13	< .001		
Interaction	-2.048	0.35	-0.37	-5.93	< .001		

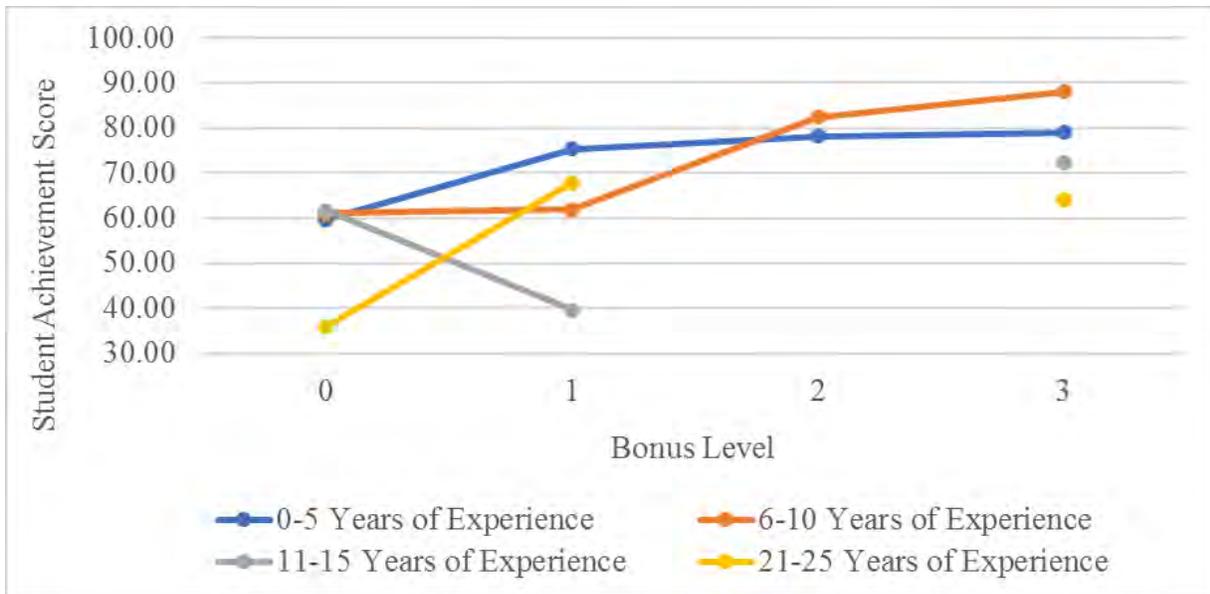
As shown in Figures 2 through 5, there were significant moderation effects for teacher years of experience, level of education, and content area, but not for teacher gender. Teachers with higher education levels had students with significantly higher scores at either end of the bonus scale. However, for teachers at a bonus level of 1, the teachers with less experience had students with higher scores than teachers with more experience. The moderation effects for years of experience and content area were much more confusing, with less discernible patterns. Overall, students in the English classes (English 1 or English 2) had higher academic achievement scores than students in the Algebra classes (Algebra 1 or Algebra 2), regardless of bonus level.

Figure 2: Moderation Effect by Education Level



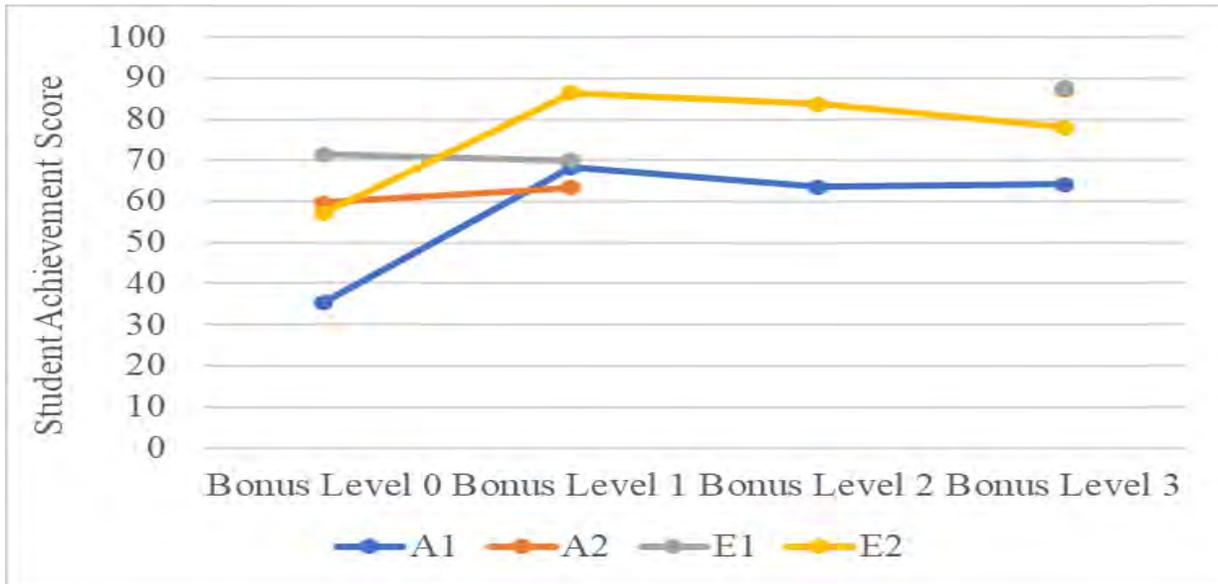
Note: Education level 1=Bachelor's degree, Education level 2=Master's degree. 2011: 0=no bonus, 1=\$237, 2=\$356, 3=\$533; 2012 and beyond: 0=no bonus, 1=\$400, 2=\$650, 3=\$900.

Figure 3: Moderation Effect of Years of Experience



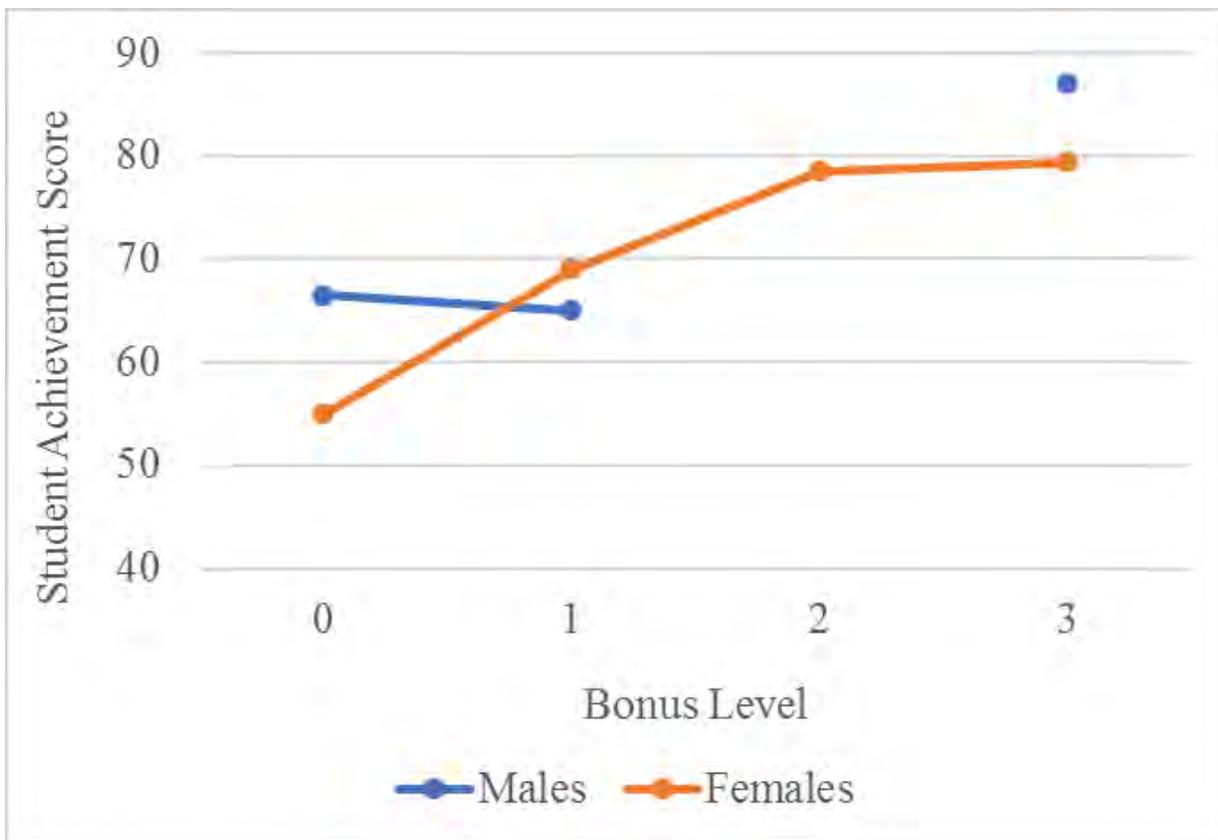
Note: 2011: 0=no bonus, 1=\$237, 2=\$356, 3=\$533; 2012 and beyond: 0=no bonus, 1=\$400, 2=\$650, 3=\$900.

Figure 4: *Moderation Effect of Content Area*



Note: A1=Algebra I, A2=Algebra 2, E1=English 1, E2=English 2. 2011: 0=no bonus, 1=\$237, 2=\$356, 3=\$533; 2012 and beyond: 0=no bonus, 1=\$400, 2=\$650, 3=\$900.

Figure 5: *No Moderation Effect by Teacher Gender*



Summary of Findings

To address the first research question, examining the relationship between teacher effectiveness levels (significantly below expectations, below expectations, at expectations, above expectations, and significantly above expectations) and teacher merit pay level in a school district, a chi-square analysis was conducted. The Chi-square test revealed a highly significant difference between expected and observed counts comparing the two categorical groups ($p < 0.001$, Table 4). Considerably, more teachers at the highest effectiveness level (significantly above expectations) also belong to higher bonus levels than what would be expected in a normal distribution. This indicates that higher teacher effectiveness may have a relationship with higher bonuses.

To address the second research question, examining the relationship between teacher merit pay levels and student achievement scores, a one-way ANOVA was conducted. Level of bonus served as the independent variable and student achievement scores as measured by the TCAP End-Of-Course state achievement tests (percent of questions that each student answered correctly) served as the continuous dependent variable. ANOVA and subsequent post-hoc pair-wise comparison indicated that student achievement scores when teachers did not receive a bonus were significantly less than any and every other bonus level (Figure 1 and Table 5). Levene's assumption test indicated a violation of data heterogeneity ($p < 0.05$). Therefore, one-way ANOVA was also calculated with Brown-Forsythe and Welch adjustments to determine if controlling for non-homogeneity affected significance levels. All 3 ANOVA tests indicated highly significant differences among bonus levels ($p < 0.001$), signifying true differences in spite of the initial assumption violation. Post-hoc comparison indicated that student achievement scores when teachers did not receive a bonus were significantly less than any and every other bonus level.

To address the third research question, a linear regression analysis with moderation was conducted to examine whether the association between bonus level and student achievement scores are moderated by teacher demographics. To create the moderation variables, teacher bonus level was multiplied by each teacher demographic characteristic to create 4 separate interaction effects (i.e., bonus level*education level, bonus level*gender, bonus level*content area, bonus level*years of experience). Separate linear regression analyses were conducted to examine each potential moderator. In each analysis, teacher bonus level, the teacher demographic characteristic, and the associated moderation variable were entered as independent variables, and student achievement score was entered as the dependent variable.

There were significant moderation effects for teacher years of experience, level of education, and content area, but not for teacher gender. Teachers with higher education levels had students with significantly higher scores at either end of the bonus scale. However, for teachers at the lowest bonus amount, the teachers with less experience had students with higher scores than teachers with more experience. The moderation effects for years of experience and content area were much more confusing, with less discernible patterns. Overall, students in the English classes (English 1 and English 2) had higher academic achievement scores than students in the Algebra classes (Algebra 1 or Algebra 2), regardless of bonus level.

Results

The results for the first research question regarding the relationship between teacher effectiveness and teacher merit pay show significant connection in general. This result fits with some of the findings in the literature review that showed increasing teacher incentives lead to higher teacher effectiveness (Goldhaber & Walch, 2012; Dee & Wyckoff, 2015; Winters, Greene, Ritter, & Marsh, 2008). The Chi-Squared test revealed a highly significant difference between expected and observed counts comparing the two categorical groups ($\chi^2 = 2302.48$ $p < .001$). Results indicated that considerably more teachers at the highest effectiveness level (significantly above expectations) also belong to higher bonus levels than what would be expected in a normal distribution.

The literature review provided support against performance pay programs because extrinsic motivators in the form of monetary compensation can decrease morale and organizational trust contributing to a lack of collaboration among teachers (Podgursky & Springer, 2007). Furthermore, external incentives in the form of rewards and punishments have a detrimental effect on intrinsic motivation (Choi, 2015; Shifrer, Turley, & Heartd, 2017). Additionally, financial incentives could affect teacher autonomy and professionalism (Shifrer, et al., 2017).

With the discrepancy between previous findings and the results of this research question, it is crucial to refer to the theoretical framework. Expectancy theory has been previously applied to merit pay. The theory lays out 3 components of motivation: valence, expectancy, and instrumentality (Northouse, 2016; Rice, Malen, Jackson, & Hoyer, 2015; Vroom, 1964). Valence, how attractive the reward is, could have played a factor in the teacher's motivation as well as expectancy, the individual's belief that their quality work will lead to their desired outcome, and instrumentality, the individual's confidence in their ability to do that work (Northouse, 2016; Rice, Malen, Jackson, & Hoyer, 2015; Vroom, 1964). However, no statistical analysis was performed to assess if one of these elements had a greater effect on a teacher's motivation than another. Therefore, this research can be used to reinforce that expectancy theory is relatable to merit pay. The results also offer a starting point for future qualitative research to see which one of these components may have had a greater stimulus than others. Goal-setting theory aligns with the results of the study as a promise of increased compensation resulted in higher student achievement.

The results of the second research question indicated a significant relationship between merit-based pay and student achievement. This result is supported by the previous literature as several studies on performance pay in the United States have shown a positive effect on measures of student achievement (Goldhaber & Walch, 2012; Dee & Wyckoff, 2015; Vigdor, 2008; Winters, Greene, Ritter, & Marsh, 2008). However, studies have also shown negative or no effects on student test scores under performance pay programs. Minnesota's Q-Comp program included by individual and school-level bonuses and had no significant effect on test scores (Choi, 2015). New York's program using school-level incentives resulted in mostly negative, but insignificant effects (Fryer, 2013). Still other programs show no significant effect on student test scores, including Chicago's TAP and Texas' Educator Excellence Grant (Glazerman & Seifullah, 2012; Springer, et al., 2009b). However, the discrepancy between these studies and the results from this investigation may be mitigated after including the Brown-Forsyth and Welch adjustment. Upon

completing the adjusted one-way ANOVAs, the results indicated high significance among all bonus levels thereby diminishing the initial assumption violation. Additionally, the post-hoc comparison stipulated that the complete absence of bonuses generated the worse results. This result justifies the goal-setting theory that when teachers are persistent in pursuing realistic goals, there is an increase in task performance, in this case, student achievement. Also, this result aligns with expectancy theory that teachers may have been motivated by a financial award to increase student achievement. There is also an opportunity, much like in the first research question, to further investigate how valence, expectancy, and instrumentality influenced the results individually.

The results of the third research question indicated there were moderating effects of demographic information for teachers' years of experience, level of education, and content area on the relationship between merit-based pay and teacher effectiveness and student achievement. There was no corresponding relationship with the variable of teacher gender. These results largely align with existing research. The literature review has shown that education and experience of teachers can influence student achievement. Hattie (2009) found that the most significant aspect of student achievement was teacher expertise. Metzler and Woessmann (2010) found that a one standard deviation increase in teacher subject knowledge in math raises student test scores by about 9% of a standard deviation. In a review of research studies on teacher knowledge and student achievement, Kleickmann et al. (2013) found that teachers' subject matter knowledge significantly increases student achievement gains. Keller, Neumann, and Fischer (2017) explained the effect teacher knowledge on student achievement because higher content knowledge challenges and supports student learning by anticipating student difficulties and adapting to student problems. However, the variable of gender and the theoretical frameworks are less clear with their relationship to the results. Like the research on merit pay as related to student achievement, the research on merit pay as related to teacher gender also has mixed results. Hendricks (2014) found no evidence that the effects of merit pay vary by gender. This is in contrast to a study that found women were more likely to receive merit pay bonuses than men (Belfield & Heywood, 2008).

Years of experience and level of education can correspond to instrumentality. Higher levels of education or more experience teaching could improve a teacher's belief that they are capable of performing the work to meet their goals, thereby increasing their motivation to earn merit pay. Future theoretical research could focus on what can influence instrumentality among teachers to further progress the theoretical framework.

Implications and Conclusion

The purpose of this study was to investigate whether merit-based pay increase teacher effectiveness and student achievement. The intent was that the results of the study might contribute to the knowledge of educational leaders regarding the usefulness in merit-based pay. Unfortunately, the study's results do not provide clear evidence. However, this study's findings still have implications for leadership. The study highlights the need for conclusive evidence in support of merit-based pay before leaders expand upon the bonus component of the compensation plan.

The ambiguity in interpreting the study's results highlights another important aspect. The results could be used in either support for or against merit-based pay depending on how they are interpreted. Thus, it is imperative that those in leadership positions look at the research with great scrutiny. It is too easy to manipulate data even unknowingly to support one's agenda. This study can serve as a reminder to always use caution when interpreting data and research and to look at results from different angles. Education is too important to not consider all of the contributing factors in reform.

In the district studied, the findings could be used to inform the use of the bonus component of the compensation plan. However, the leaders of the school district must look at the results in detail. While the results indicate merit-based pay had a significant and positive effect on student achievement and teacher effectiveness scores, it also showed a fair amount of inconsistency and lack of conclusive data. Therefore, district leaders should use caution when interpreting the results. Due to ambiguity, the district may want to search for other components of differentiated pay plans to meet the desired goal of increasing teacher effectiveness and student achievement. Additional studies could provide clearer evidence of whether the intended consequence of paying teachers for performance is successful or not.

By investigating the changes in student achievement in high school Math and English, district stakeholders can determine if rewarding teachers for student test scores is an effective component in district strategic compensation plans. By including merit pay, more effective teachers will be retained, and student achievement will continue to increase. However, if teachers perceive the effort not worth the reward, student test scores could suffer, and teachers could seek other employment. As stated previously, the research on the effects of merit pay especially on student achievement has been inconclusive in the past and focused mainly on larger school systems.

Policymakers at both the local and state level must be aware of the research on how merit pay relates to student achievement and teacher effectiveness. So far, the results of the research remain inconclusive. This suggests that more research is needed as well as an exploration of different ways of conducting education reform in reference to merit-based pay.

Leaders in education are faced with difficult challenges. Those with power must be diligent to serve the greater good and not one's own agenda especially when it comes to resource allocation around education. The current study produced results that were supportive of merit pay. Yet these results were far from conclusive as many were often confounding and conditional upon closer inspection. Additionally, the conflicting nature of existing literature further supports the mixed signals that were yielded from this study. The results of this study ultimately contributed a modest addition to existing knowledge of the phenomena, but provide a distinct jumping off point for future research. The problem of education reform and increasing student achievement remains complicated and unresolved.

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–136.
- Alexander, N. A., Jang, S. T., & Kankane, S. (2017). The performance cycle: The association between student achievement and state policies tying together teacher performance, student achievement, and accountability. *American Journal of Education*, 123(3), 413–446.
- American Statistical Association. (2014). *ASA Statement on using value-added models for educational assessment*. Retrieved from http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf
- Archer, J., Kerr, K., & Pianta, R. (2014). Why measure effective teaching. In T. Kane, A. Kerri, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 1–5). San Francisco, CA: John Wiley.
- Ballou, D., Springer, M. G., McCaffrey, D. F., Lockwood, J. R., Stecher, B. M., Hamilton, L., & Pepper, M. (2012). POINT/CounterPOINT: The view from the trenches of education policy research. *Education Finance and Policy*, 7(2), 170–202.
- Belfield, C. R., & Heywood, J. S. (2008). Performance pay for teachers: Determinants and consequences. *Economics of Education Review*, 27(3), 243–252.
- Bowen, D. H., & Mills, J. N. (2017). Changing the education workforce? The relationships among teacher quality, motivation, and performance pay. *Teachers College Record*, 119(4), 1–32.
- Choi, W. S. (2015). *The effect of alternative compensation programs on teacher retention and student achievement: The case of Q comp in Minnesota* (University of Minnesota). Retrieved from <https://search-proquest-com.ezproxy.memphis.edu/pqdtglobal/docview/1672175072/fulltextPDF/E3CF3810A5814102PQ/44?accountid=14582>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed method approaches*. Thousand Oaks, CA: Sage.
- Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132–137.
- Dee, T. S., & Jacob, B. A. (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418–446.
- Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 35(2), 252–279.
- Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34(2), 267–297.
- Dee, T. S., & Wyckoff, J. (2017). A lasting impact: High-stakes teacher evaluations drive student success in Washington, D.C. *Education Next*, 17, 58–66.
- Deming, D. J., & Figlio, D. (2016). Accountability in US education: Applying lessons from K-12 experience to higher education. *The Journal of Economic Perspectives*, 30(6), 33–55.
- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, 34, 12–25.
- Dubnick, M. J. (2006). Orders of accountability. Paper present at the World Ethics Forum, Oxford, UK. Retrieved from <http://pubpages.unh.edu/dubnick/papers/2006/oxford2006.pdf>
- Eberts, R., Hollenbeck, K., & Stone, J. (2002). Teacher performance incentives and student outcomes. *The Journal of Human Resources*, 37(4), 913. Retrieved from <https://doi.org/10.2307/3069621>
- Editorial Projects in Education Research Center. (2004). Issues A-Z: Accountability. *Education Week*. Retrieved from <https://www.edweek.org/ew/issues/accountability/index.html>
- Fernandez, C. (2014). Knowledge base for teaching and pedagogical content knowledge (PCK): Some useful models and implications for teachers' training. *Problems of Education in the 21st Century*, 60, 79–100.
- Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, 31(2), 373–407.
- Glazerman, S., & Seifullah, A. (2012). An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years. Final Report. *Mathematica Policy Research, Inc.*
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, 2(1), 50–55.

- 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.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2).
- Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education*, 112(2), 209–238.
- Hatry, H. P., Greiner, J. M., & Ashford, B. G. (1994). *Issues and case studies in teacher incentive pay plans* (2nd ed.). Washington, DC: Urban Institute Press.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hendricks, M. D. (2014). Does it pay to pay teachers more? Evidence from Texas. *Journal of Public Economics*, 109, 50–63.
- Heneman III, H. G. (1998). Assessment of the motivational reactions of teachers to a school-based performance award program. *Journal of Personnel Evaluation in Education*, 12(1), 43–59.
- Holloway, J., Sørensen, T. B., & Verger, A. (2017). Global perspectives on high-stakes teacher accountability policies: An introduction. *Education Policy Analysis Archives*, 25, 85.
- Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies*, 40(4), 417–456.
- Iasevoli, B. (2017). Merit pay for teachers can lead to higher test scores for students, a study finds - Teacher Beat - Education Week. Retrieved April 3, 2019, from http://blogs.edweek.org/edweek/teacherbeat/2017/04/_meritpaystudy.html
- Jennings, J., & Lauen, D. (2016). Accountability, inequality, and achievement: The effects of the No Child Left Behind act on multiple measures of student learning. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(5), 220–241.
- Kaimal, G., & Jordan, W. J. (2016). Do incentive-based programs improve teacher quality and student achievement? An analysis of implementation in 12 urban charter schools. *Teachers College Record*, 118(7), 1–34.
- Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, 54(5), 586–614.
- Kelley, C., Heneman, H., & Milanowski, A. (2002). Teacher motivation and school-based performance awards. *Educational Administration Quarterly*, 38, 372–401.
- Kellough, J. E., & Lu, H. (1993). The paradox of merit pay in the public sector. *Review of Public Personnel Administration*, 13(2), 45–64.
- Kimball, S. M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education*, 16(4), 241–268.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106.
- Lauen, D. L. (2013). Jumping at the chance: The effects of accountability incentives on student achievement. *Journal of Research on Educational Effectiveness*, 6(2), 93–113.
- Lavigne, A. L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record*, 116(1), 1–29.
- Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science*, 15(5), 265–268.
- Loewenber Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- McMillan, J. H. (2016). *Fundamentals of educational research* (7th ed). Pearson Education.
- Metzler, J., & Woessmann, L. (2010). The impact of teacher subject knowledge on student achievement: Evidence from within- teacher within-student variation. *Journal of Development Economics*, 99(2), 486–496.
- Milanowski, A. T., & Heneman, H. G. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193–212.
- Mintrop, R., Ordenes, M., Coghlan, E., Pryor, L., & Madero, C. (2018). Teacher evaluation, pay for performance, and learning around instruction: Between dissonant incentives and resonant procedures. *Educational Administration Quarterly*, 54(1), 3–46.

- National Assessment of Educational Progress (NAEP). (2009). *Nation's report card*. Retrieved from www.nationsreportcard.gov
- National Center for Education Statistics. (2003). Total and current expenditures per pupil in fall enrollment in public elementary and secondary education, 2002-2003. Retrieved May 25, 2018, from https://nces.ed.gov/programs/digest/d16/tables/dt16_236.55.asp
- Northouse, P. G. (2016). *Leadership: Theory and practice* (7th ed.). Sage Publications.
- Odden, A., Kelley, C., Heneman, H., & Milanowski, A. (2001). *Enhancing teacher quality through knowledge and skills-based pay*. Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from http://www.cpre.org/sites/default/files/policybrief/875_rb34.pdf
- Office of Innovation and Improvement. (n.d.). Teacher incentive fund. Retrieved from <https://innovation.ed.gov/what-we-do/teacher-quality/teacher-incentive-fund/>
- Pham, L. D., Nguyen, T. D., & Springer, M. G. (2017, June). Teacher merit pay and student test scores: A meta-analysis. In *Association for Education Finance and Policy Annual Meeting, Washington, DC*.
- PISA. (2015). Key findings from PISA 2015 for the United States. In *OECD Better Policies for Better Lives*. Retrieved from <https://www.oecd.org/pisa/PISA-2015-United-States.pdf>
- Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, 26(4), 909–949.
- Provasnik, S., Malley, L., Stephens, M., Landeros, K., Perkins, R., & Tang, J. (2016). *Highlights from TIMSS and TIMSS advanced 2015: Mathematics and science achievement of U.S. students in grades 4 and 8 and in advanced courses at the end of high school in an international context (NCES 2017-002)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Range, B. G., Young, S., & Hvidston, D. (2013). Teacher perceptions about observation conferences: What do teachers think about their formative supervision in one US school district? *School Leadership and Management*, 33(1), 61-77.
- Rice, J. K., Malen, B., Jackson, C., & Hoyer, K. M. (2015). Time to pay up: Analyzing the motivational potential of financial awards in a tif program. *Educational Evaluation and Policy Analysis*, 37(1), 29–49.
- Rice, J., Malen, B., Jackson, C., & Hoyer, K. M. (2017). Administrator responses to financial incentives: Insights from a tif program. *Leadership and Policy in Schools*, 16(3), 475–501.
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2007). *Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure*. Cambridge, MA: National Bureau of Economic Research.
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020–1049.
- Shifrer, D., Turley, R. L., & Heard, H. (2017). Do teacher financial awards improve teacher retention and student achievement in an urban disadvantaged school district? *American Educational Research Journal*, 54(6), 1117–1153.
- Shulman, L. S. (1986). Those who understand: A conception of teacher knowledge. *Educational Researcher*, 15(2), 4–14.
- Smith, W. C. (2016). National testing policies and educator based testing for accountability: The role of selection in student achievement. *OECD Journal: Economic Studies*, 2016, 131–148.
- Spring, J. (2016). *American Education*. New York, NY: Routledge.
- Springer, M. G., Lewis, J. L., Podgursky, M., Ehlert, M. W., Gronberg, T. J., Hamilton, L. S., ... Peng, A. (2009a). *Governor's educator excellence grant (GEEG) program: Year three evaluation report*. Retrieved from https://my.vanderbilt.edu/performanceincentives/files/2012/10/200908_SpringerEtAl_GEEG_Year3.pdf
- Springer, M. G., Lewis, J. L., Podgursky, M., Ehlert, M. W., Gronberg, T. J., Hamilton, L. S., ... Peng, A. (2009b). Texas educator excellence grant (TEEG) program: Year three evaluation report. 549. *National Center on Performance Incentives*.
- Stiggins, R., & Duke, D. (1988). *The case for commitment to teacher growth: Research on teacher evaluation*. Albany, NY: State University of New York Press.

- Tchoahanov, M. A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics*, 76(2), 141–164.
- Tennessee Department of Education (TDOE). (2011). *System achievement level summary report*.
- Tennessee Department of Education. (2013). *Tennessee educator acceleration model (TEAM) TEAM evaluator handbook*. Retrieved from http://team-tn.org/wp-content/uploads/2013/08/TEAM-Teacher-Evaluator-Handbook-2017-18_Add-Gifted-Doc.pdf
- Tennessee Department of Education (TDOE). (2017). Tennessee state report card. Retrieved from <https://reportcard.tnk12.gov/>
- Tennessee State Board of Education (2018a). High school policy 2.103., Pub. L. No. 2.103, High School Policy.
- Tennessee State Board of Education (2018b). Teacher and administrator evaluation policy 5.201., Pub. L. No. 5.201, Tennessee State Board of Education Rules and Policies 25.
- Toh, T. L. (2017). On Singapore prospective secondary school teachers' mathematical content knowledge. *International Journal for Mathematics Teaching and Learning*, 18, 25–40.
- U.S. Department of Education. (2012). *Teacher incentive fund: First implementation report, 2006 and 2007 grantees*. Retrieved from <https://www2.ed.gov/rschstat/eval/teaching/tif/report.pdf>
- Vigdor, J. L. (2008). Teacher salary bonuses in North Carolina. Working Paper 2008-03. *National Center on Performance Incentives*.
- Vroom, V. H. (1964). *Work and motivation*. (Vol. 54). New York, NY: Wiley.
- Winters, M. (2011). Measuring teacher effectiveness: Credentials unrelated to student achievement. Issue Brief No. 10. *Manhattan Institute for Policy Research*.
- Winters, M., Greene, J., Ritter, G., & Marsh, R. (2008). The effect of performance pay in Little Rock, Arkansas on student achievement. *National Center on Performance Incentives*.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57–67.
- Yuan, K., Le, V. N., McCaffrey, D. F., Marsh, J. A., Hamilton, L. S., Stecher, B. M., & Springer, M. G. (2013). Incentive pay programs do not affect teacher motivation or reported practices: Results from three randomized studies. *Educational Evaluation and Policy Analysis*, 35(1), 3-22.