WORTH THE FIGHT? TEACHERS' COLLECTIVE BARGAINING AGREEMENTS AND STUDENT OUTCOMES

Adam Kirk Edgerton

A DISSERTATION

In

Education

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

_____

Manuel S. González Canché, Associate Professor of Education

Graduate Group Chairperson

_____

Matthew Hartley, Professor of Education

Dissertation Committee:

Manuel S. González Canché, Associate Professor of Higher Education

Rand Quinn, Associate Professor of Teaching, Learning, and Leadership

Laura M. Desimone, Director of Research, University of Delaware

WORTH THE FIGHT? TEACHER COLLECTIVE BARGAINING AGREEMENTS

AND STUDENT ACHIEVEMENT

DEDICATION

This dissertation is dedicated to my mother. A lifelong nurse, she reminds me that all that I do should be in the service of others.

ACKNOWLEDGMENTS

First, I would like to acknowledge the best advisors anyone could ask for, Laura and Manuel, as well as Rand, the third member of my committee. This process has been smooth, productive, and constructive because of them. I could not ask for better mentors. Laura has been my rock throughout graduate school.

I also thank my father, who inspired me with a love of political science at an early age and taught me to reach for my maximum potential. I thank Travis, my fiancé, who met me during a particularly low moment in my graduate career after a difficult first year in Philadelphia. He has seen me through the trials and tribulations of the job market. I also thank my two best friends, Haisheng Yang and Josh Coleman, without whom I would not have survived these past four years. I am grateful to have come to Philadelphia knowing nothing and no one and to have gained so much love and support.

Finally, thank you to those who aided in the collection of the original collective bargaining agreements (Jessica Barnett of the Commonwealth Foundation), the graduate students who assisted with the coding of the contracts (Abby Quirk, Joao Gomes, Michalea Ward, and Marissa Gonzalez), and Andy Porter, who generously covered the stipends for these graduate students and wrote dozens of letters of recommendations on my behalf.

ABSTRACT

WORTH THE FIGHT? TEACHER COLLECTIVE BARGAINING AGREEMENTS AND STUDENT OUTCOMES

Adam Kirk Edgerton

Manuel S. González Canché

Education reform rhetoric frequently pits the vested interests of teachers' unions against those of students and families. To test whether union restrictions are related to student learning, I analyze a unique database of contractual items for the 2016-2017 school year across all 499 Pennsylvania school districts in order to examine a) variation, b) partisan political predictors, and c) relationships to student achievement and graduation rates. I also examine changes in 105 contracts that occurred during the 2015-2016 school year.

I depict variation among items using GIS mapping. I use OLS regression, probit regression, and spatial autoregression to examine relationships between contract features and student proficiency and graduation rates. I also use propensity score weighting with generalized boosted models (GBM).

After controlling for spatial dependence and district demographics, I find a significant negative relationship between the percentage of registered Republicans in a district and bonuses for teacher graduate credentials. I find a significant and positive relationship between Republican registered voters and math and science proficiency. This

relationship diminishes in magnitude for ELA proficiency. I also find a significant positive relationship between average years of teaching experience and ELA proficiency in grades 3-8.

Using GBM, I find significant positive estimates (+2%) of teacher qualification indicators on students' math achievement in grades 3-8, and a significant positive estimate (+2%) between harsh consequences for ELA teachers and student proficiency. I also find a significant positive estimate between higher teacher pay and biology proficiency (+4% for historically disadvantaged students), as well as a significant negative estimate of graduate credential bonuses on graduation rates (-6%).

These correlational results suggest that subject-area and grade-level differentiation in contracts – such as higher wages for STEM teachers – might be beneficial. The most effective STEM teachers might be seeking out positions in the best-paying districts with the strongest contracts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION

Tearing up the Contract in Lawrence, Massachusetts

When the Supreme Court ruled in 2018 in *Janus v. AFSCME* that teachers who opt out of joining their union no longer have to pay mandatory fair share fees, I felt ambivalent. As a novice high school English teacher, I had primarily negative interactions with my local union. I saw my union representative in our building only when my colleagues conducted themselves poorly and faced serious disciplinary action. Whenever I watched our union defend our most problematic teachers, I felt resentful, knowing that I was helping to pay for their arbitration. At times, it seemed as though the more junior members of our department put in twice as much work for less than half of a veteran teacher's salary. As new teachers, we rarely had contact with the local union, which was run by the more senior members. Participation in union leadership or meetings was not encouraged nor cultivated among the new staff, and power seemed concentrated among a small group of veterans. Our superintendent, meanwhile, was in the process of being convicted for corruption and embezzlement (Schworm, 2012). My first impression of teachers' unions, as well as local district politics, was far from positive.

Yet one year later, I witnessed the other policy extreme, when we lost many of our collective bargaining rights. Under Massachusetts's 2010 Achievement Gap Act, a state-appointed receiver was able to "request that the school committee and any union bargain or reopen the bargaining of the relevant collective bargaining agreement to facilitate such [student] achievement," and to "limit, suspend, or change" the collective bargaining agreement (CBA) (M.G.L. Ch 69, Section 1J). The passage of this legislation

was somewhat surprising in a heavily unionized and heavily Democratic state like Massachusetts, but it was in keeping with neoliberal, bipartisan education reforms of the era (Au, 2016). These reforms located the source of achievement problems inside of individual classrooms, and it placed the burden of accountability on teachers and students, not elected representatives and district administrators. Under the 2010 law – passed specifically with our district in mind – the state effectively tore up our district's collecting bargaining agreement (CBA).

There are now several turnaround districts in Massachusetts, where the local teachers' union is neutralized on matters that it once tightly controlled. Turnaround is a dramatic shift in governance, reallocating power and deciding who can and cannot participate in policy debates (Dahill-Brown, 2019; Morel, 2018). Across the nation, the weakening of collective bargaining has become both a rhetorical and a legal cornerstone to enact education reforms, reforms ostensibly done to raise student achievement.

There now exist causal studies that examine the totality of the turnaround process in Lawrence, which answer the question of whether student achievement increased as a result. Ultimately, whole-district turnaround in the Lawrence Public Schools significantly raised achievement at the elementary grades, but achievement effects did not manifest at the high school where I taught (Schueler, Goodman, & Deming, 2017). Said the lead author of this study, Dr. Beth Schueler, "Lawrence is therefore a rare positive example of politically viable state-led district-wide improvement that provides lessons for state-level policymakers on how to select districts for turnaround, and for district-level leaders on navigating the thorny politics of school system improvement" (Schueler, 2018). As a

teacher during this study's time frame, I am more skeptical of these positive claims, which are based on statistically significant test score gains. I question whether the elimination of the CBA was truly necessary to achieve reform – or, more narrowly defined, to increase student achievement. As part of making sense of my professional experience in Lawrence, this dissertation focuses on discerning the relationships between CBAs and student achievement in a state that is more politically divided than Massachusetts, and one that is rarely studied in the CBA literature – Pennsylvania.

Though there are many components of whole-district turnaround (see Supovitz, 2006), the first step for the Lawrence Public Schools necessitated the nullification of the CBA in order to reallocate teachers and fire ineffective ones. According to Papay and Hannon (2018), turnaround in Lawrence also increased teacher quality, when measured as student growth in standardized tests in mathematics and English Language Arts (ELA). In the case of Lawrence, "gap closure" policies (see Gutiérrez & Dixon-Román, 2011) replaced veteran faculty of color with young White teachers, as has been the case nationally (Ingersoll & Merill, 2017; Yarnell & Bohrnstedt, 2017). Teach for America also played a pivotal role in Lawrence, which has been shown to Whiten the teaching force in many districts (Trujillo, Scott, & Rivera, 2017). These new teachers were indeed more effective at raising student achievement (Papay & Hannon, 2018). But the focus on teacher recruitment in Lawrence, instead of the retention and development of existing staff, mirrors a larger national trend. States with weaker unions use extra funds to increase teacher recruitment, while states with stronger unions invest in the existing workforce (Brunner, Hyman, & Ju, 2019).

In this dissertation, I do not study the impact of teachers' collective bargaining as a whole on student outcomes – a question for which large national datasets are more suitable (e.g., Brunner, Hyman, & Ju, 2019; Lovenheim & Willén, 2019). Instead, I consider which specific CBA items might be beneficial for student outcomes, and which might be beneficial only to teachers. Much political capital is spent on attempting to change these items, as I witnessed firsthand. Throughout the state-led turnaround process, I questioned whether my local union acted in the best interests of our students, or at the expense of our students. I now ask whether this dichotomy is false – perhaps teacher interests *are* student interests. Happier teachers may make for happier, higher-achieving students (Darling-Hammond, 2003). Was the nullification of the CBA a wellspring from which a package of reforms could flow? Or, could reforms have been achieved without sacrificing collective bargaining rights?

In considering these questions, I assert that the CBA itself, the document, shoulders much of the blame for poor student achievement. When politicians reference "obstructionist unions," they are implicitly referring to the CBA that codifies much of the union's influence (Moe, 2015). Reformers see the CBA as the first and largest obstacle to policy change. CBAs, a proxy for the strength of the local union (Strunk, Cowen, Goldhaber, Marianno, Kilbride, & Theobald, 2018), consume much political attention despite a relative lack of empirical evidence as to their importance. Throughout this work, I dig into the intricacies of CBAs, parsing out items from seniority rights to maximum class sizes. Evidence in other states suggests that there are small negative effects for some of these items, such as class size and compensation restrictions, though

not for the contract as a whole (Marianno & Strunk, 2018). I discuss each of these studies in detail in the literature review contained in Chapter 2.

Ultimately, I provide an empirical foundation for understanding the landscape of 499 collective bargaining agreements across Pennsylvania. I read and code all of these CBAs to answer descriptive, predictive, and quasi-experimental questions. This work remains important and timely. CBAs can define much of the working conditions for teachers for the better by ensuring that teachers have mandatory prep time, to give one example. These improved working conditions may lead to improved outcomes for students (Darling-Hammond, 2003). CBAs also shape much of a district's policy (Strunk & Grissom, 2010), which affects not only the professional satisfaction of teachers but also student achievement, particularly in high-need schools (Johnson, Kraft, & Papay, 2012).

Why Pennsylvania?

Pennsylvania has 499 school districts with enrollment ranging from 199 students to 140,000, and it has unusually prescriptive laws related to CBAs. There are 168 rural, 75 town, 236 suburban, and 20 urban districts in Pennsylvania, according to the National Center for Education Statistics (2018). One of these districts has no public schools and is thus excluded from the study. The prescriptiveness of the state policy environment makes studying Pennsylvania an important contribution to the limited literature available. The absence of a state east of Ohio from the existing literature makes Pennsylvania a particularly important site for research, as differences in education governance are "very often underappreciated" (Dahill-Brown, 2019, p. 37). With fewer differences among

Pennsylvania's districts compared to the other previously studied states, I expect to find stronger relationships among those items that are variable.

Pennsylvania has had several years of policy stability as a result of Pennsylvania's Act 82, passed in 2012, which amended the state School Code to include many popular reform policies, such as holding teachers accountable for student achievement through centralized, standardized evaluation practices. Unlike the previous four states where CBAs have been studied (see: Strunk et al., 2018), Pennsylvania mandates items such as the inclusion of student achievement measures. Such specificity enabled the state to develop a winning Race to the Top (RttT) application, including "a new multiple-measure rating system to evaluate teachers and principals, specified policies for teacher dismissal, and mandated participation in PD for all teachers rated as low performing" (Stecher, Holtzman, Garet, Hamilton, Engberg, Steiner, Robyn, Baird, Gutierrez, Peet, de los Reyes, Fronberg, Weinberger, Hunter, & Chambers, 2018, p. 31). RttT, a $4.3 billion grant competition during the Obama administration, effectively incentivized the adoption of all of these policies (Bleiberg & Harbatkin, 2018; McGuin, 2012).

Currently, the Pennsylvania School Code dramatically restricts the range of contractual items up for negotiation. 99 percent of districts now use the state-provided form to evaluate teachers, according to my correspondence with an official at the Pennsylvania Department of Education (PDE). Act 82 requires the use of student achievement measures, which must represent 50% of a teacher's evaluation score. 20% of the 50% can come from elective data, "including measures of student achievement that are locally developed and selected by the school district from a list approved by the

department and published in the Pennsylvania Bulletin by June 30 of each year." The remainder (30%) must be valued-added scores based on the relevant state standardized test (either the reading or mathematics Pennsylvania System of School Assessment [PSSA] in grades 3-8, or the Keystone exam for high school students in Algebra I, Algebra II, Geometry, Literature, English Composition, Biology, Chemistry, U.S. History, World History, and Civics and Government) (PDE, 2018).

Pennsylvania also ranks highly in terms of the specificity it provides and the resources it offers to help districts evaluate teachers in a uniform fashion (The Center on Standards, Alignment, Instruction, and Learning [C-SAIL], 2019). In short, districts have little wiggle room by which to substantially alter their teacher evaluation systems; the remaining 50% of a teacher's score comes from classroom observations. Still, a teacher's students could perform relatively poorly on the state test and still receive a rating of Proficient, and teachers can be awarded tenure regardless of effectiveness (PDE, 2018).

There is also leeway afforded for establishing consequences for evaluation scores, class size maximums, National Board certification bonuses, and other items listed in Table 1. This table represents all of the variable items I have collected from the coding of CBAs. I discuss this coding process in more depth in Chapter 3 (Methods).

Table 1

Unweighted Averages of Pennsylvania CBA Characteristics and High-Profile Provisions

Panel A: Context, union dues, and compensation	**Average**
1. Teacher salary	$63,857
2. CBA page length	48 pages
3. Members maintain union membership or pay fair share/service	0.812
4. Bonus for "Hard to Recruit" teachers	0.012
5. Bonus for Longevity	0.218
6. Bonus for National Board Certification	0.130
7. Number of rows in salary schedule	17 rows
8. Number of columns in salary schedule	6 columns
Panel B: Class size	
9. CBA includes a maximum class size for fourth-grade classrooms	0.050
10. CBA includes a maximum class size for eighth-grade classrooms	0.050
11. CBA includes a maximum class size for ninth- to 12th-grade	0.048
12. District takes action if class size ceiling is exceeded	0.040
Panel C: Teacher evaluation	
13. CBA **does not** say teachers are evaluated on student achievement	0.980
14. Bonus for multiple years of satisfactory ratings	0.004
15. Permanent/tenured members can use alternative process	0.034
16. CBA defines consequences of negative evaluations	0.136
17. CBA allows for teachers to rebut or appeal negative evaluations	0.168
18. Mandates a professional growth plan for teachers rated ineffective	0.138
Panel D: Layoffs	
19. CBA specifies that layoffs must occur in reverse order of seniority	0.431
20. CBA provides recall rights after layoffs	0.325
21. CBA specifies how reemployment offers are made after layoffs	0.244
22. CBA mandates reemployment offers in seniority order	0.244
Panel E: School day	
23. CBA includes mandated preparation time for fourth-grade	0.754
24. Average preparation time for fourth-grade classrooms	160 minutes
25. CBA includes mandated preparation time for eighth-grade	0.745
26. Average preparation time for eighth-grade classrooms	162 minutes
27. CBA includes mandated preparation time for ninth- to 12th-grade	0.748
28. Average preparation time for ninth- to 12th-grade classrooms	164 minutes

29. There is a set amount of time teachers must work per day	0.900
Panel F: Transfers and vacancies	
30. Teacher with most seniority fills position if two or more apply	0.166
31. Seniority is the deciding factor for involuntary transfers	0.206
32. CBA specifies order that district can consider new employees for	0.236
33. CBA requires that district posts all certificated vacancies	0.802

For binary items above, 0 = no, and 1 = yes.

Recently, Pennsylvania has made significant changes to how it holds schools accountable for student achievement. While relatively stringent accountability remains in place for teachers (see: Wei (2012) for a definition of "stringent accountability"), there is less stringent accountability for districts under the state's approved Every Student Succeeds Act (ESSA) plan. ESSA requires individual states to submit accountability plans to the federal Department of Education for approval, all of which have now been approved (Achieve, 2018). Pennsylvania is now abandoning summative scores for school and district accountability and embracing a multiple measures dashboard, similar to that of California (Edgerton, 2019a, 2019b; Elgart, 2016). The state has set "significantly lower goals for certain subgroups of students" and its "lack of a summative rating for schools will present significant challenges for parents seeking to quickly and easily understand the performance of their child's school" (Bellwether Education Partners, 2018, p. 2). It remains to be seen whether parents will like or dislike this new reporting system, though opinion polling in California suggests that parents prefer more complex forms of school accountability with well-populated data dashboards (PACE, 2018).

Looking at a national database of policy change, Pennsylvania is rare to make such a dramatic shift from the No Child Left Behind (NCLB) era (Pak, Edgerton, Desimone, & Song, 2019), suggesting there have been ground-level political shifts that

have yet to be examined. The state has significantly weakened its accountability systems and no longer holds districts accountable for student subgroup performance (Bellwether Education Partners, 2018).

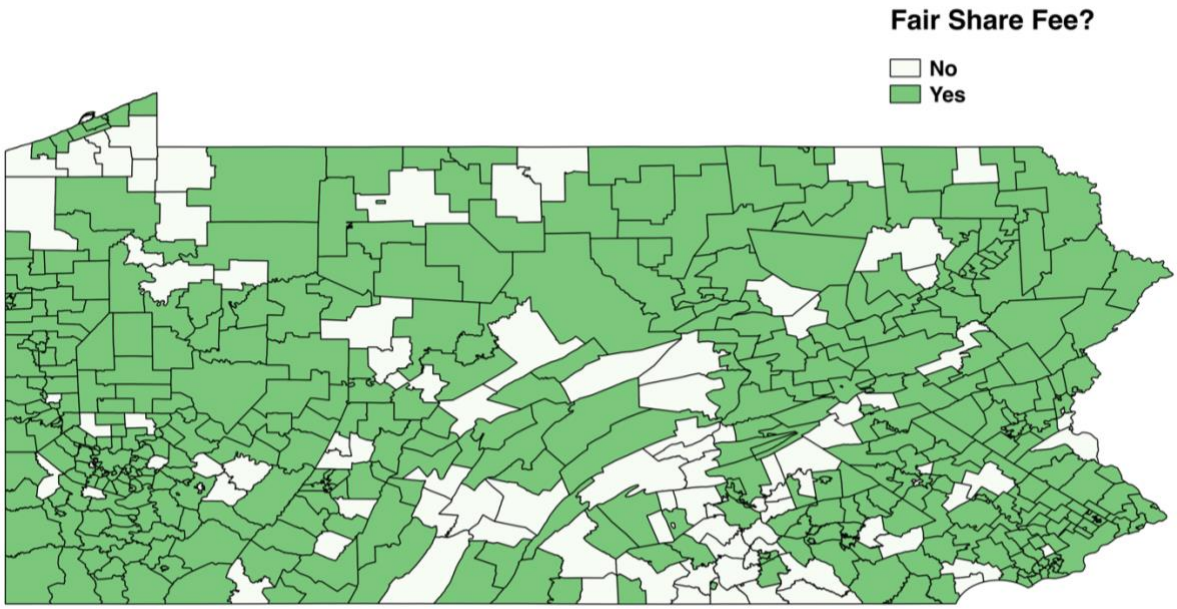
These very recent policy changes, and the ramifications of the *Janus* decision, will take years to surface in CBAs. But this dissertation provides an important foundation for understanding how contracts vary when the state education agency (SEA) wields significant control over CBAs. The findings are generalizable to states with similar policy contexts. Additionally, Pennsylvania is unique politically considering its strong history of unionization combined with an unexpected turn towards the Republican presidential candidate in 2016. The state is also the site of one of the few prior studies of school board elections; Meredith and Grissom (2010) found that negative evaluations of schools increased the competitiveness of school board elections. These political and policy characteristics make studying Pennsylvania an important step towards understanding the predictors that influence CBAs as a whole, individual CBA provisions, and the effect of these provisions on students.

Research Questions

The timing of this research is particularly important, as the recent Supreme Court decision in *Janus v. AFCSME* now prohibits the collection of fair share representation fees, which are currently required in more than 81% of school districts (Barnett, 2018). Figure 1 depicts all of the districts where fair share fees are required as of the 2016-2017 school year.

Figure 1

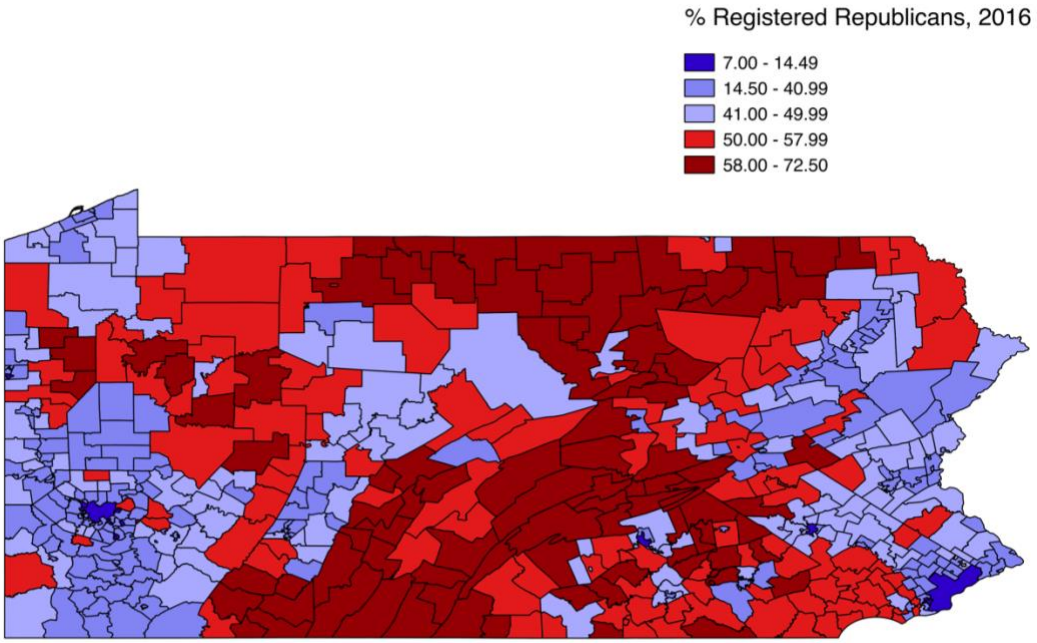
Fair Share Districts



As a result of *Janus*, all teachers in these districts will no longer required to pay these fees if they opt-out of the union, and they must now actively opt-in instead of experiencing an automatic paycheck deduction. It is important to capture the relationship between CBAs and student achievement at what may now be the height of union power before they lose a majority of their members (according to projections from Marianno and Strunk (2018)). This drop-off could be even more precipitous in areas with strong Republican school boards and, by extension, more Republican voters, as education policy issues have become more sharply partisan over the last decade (Cheng, Henderson, Peterson, & West, 2018). Figure 2 depicts the variation in registered Republicans across the state by school district, obtained from my analysis of publicly available voter registration files (Pennsylvania Department of State, 2018).

Figure 2

Percentage of Registered Republicans by School District as of 2016



In addition to researching the effects of these provisions, I examine the factors that predict the provisions within CBAs. This latter point is important given that no existing district-level quantitative research considers how the political party affiliation of school board members and CBA negotiators, as well as their constituents, relates to which CBA provisions are adopted. In the current hyper-partisan political climate, we need to better understand which issues are most contentious to highly-engaged partisans, as these political battles – not "objective" measures of school performance – often decide the configuration of public schools (Garda, 2011; Schneider, 2017).

Through a thorough reading and coding of every contract in effect during the 2016-2017 school year, I describe CBA items that remain unmandated and variable across Pennsylvania. I isolate significant predictors, in particular the partisanship of the

electorate and the school board, for the content of these CBAs, as well as estimates on student achievement. I ask and answer the following research questions:

1. To what extent do CBA items vary across districts within the state of Pennsylvania?
2. To what extent do sociopolitical, demographic, financial, and geographic spillover factors predict the presence of classroom and teacher qualification items?
3. To what extent are CBAs related to student achievement?
    a. Are these relationships different for historically disadvantaged students?
    b. Are these relationships different for individual items? For classroom and teacher qualification provisions?
4. To what extent are CBAs related to graduation rates?
    a. Are these relationships different for males, females, and Whites?
    b. Are these relationships different for individual items? For classroom and teacher qualification provisions?

For research question 4, I am only able to analyze Whites as a sub-population, as other racial and ethnic groups are not of sufficient size to be reported in the more than 100 small rural Pennsylvania school districts. After reviewing the available literature in the next chapter, I devote a chapter of this dissertation to each of the research questions above.

CHAPTER 2: LITERATURE REVIEW

This chapter describes the history of public sector collective bargaining in the United States before turning to teachers' unions more specifically. I then discuss the notion that teachers' unions represent their own vested interests (Moe, 2015), how effective they are at protecting their interests, and the relationships between collective bargaining and student outcomes in states that have been previously studied. I conclude with a discussion of Pennsylvania's politics, its collective bargaining history, and its school board governance in order to develop a theoretical framework for hypothesizing significant predictors of CBA items as well as estimates on student achievement.

Collective Bargaining and Teachers' Unions in the United States

Collective bargaining in the United States began with the passage of the National Labor Relations Act in 1935. But the first public sector collective bargaining unit did not emerge until 1959, when Wisconsin passed the first municipal relation labor law (Compa, 2014; Dismuke, 2013). Though public sector collective bargaining did not pick up steam until the 1960's, the sector is now described as a "last redoubt" thanks to huge declines in the private sector unionization over the past several decades (Scribner, 2015, p. 552). 36% percent of public employees are unionized compared to only 7% in the private sector. 31 states permit some level of public sector collective bargaining; it is banned in 19 others (Compa, 2014).

The United States was one of the last industrialized nations to implement public sector collective bargaining, but since the 1960's and 70's, teachers' unions have played an outsized role in education. They are the largest and most organized group of

stakeholders considering the relative lack of political organization among parents and students except in certain affluent and large urban communities (Moe & Wiborg, 2017). In Pennsylvania, only nine isolated school board members per district are expected to represent all of these varied interests, compared to local unions backed by one of two large national entities – either the American Federation of Teachers (AFT) or the National Education Association (NEA).

Considering their size and scope, teachers' unions are "major political forces at all levels of American government when official decisions are made about the policies, organization, funding, and reform of the public schools" (Moe & Wiborg, 2017, p. 4). They remain a potent political force despite being on the defensive in recent decades in both developed and less-developed nations. Across the globe, teachers' unions have resisted an agenda often described as neoliberal, which emphasizes a degree of increased accountability, decentralization, and school choice (p. 3; see also Cohen, Spillane, & Peurach, 2018). With few exceptions, teachers' unions have resisted this agenda of reform in favor of what Moe (2015) describes as their own vested interests. Unions prefer the status quo with existing benefits instead of policy change, and this preference has been demonstrated across the literature and on a national scale (Finger, 2017; Flavin & Hartney, 2015; Moe, 2015). Moe (2017) groups the vested interests of teachers' unions into five broad categories: salary rules, transfer rules, layoff rules, evaluation rules, and dismissal rules. I discuss the existing literature on the relationships and the effects of these policies on student achievement in a later section.

Broadly speaking, these vested interests are more pragmatic than they are ideological (Moe & Wiborg, 2017). While unions do typically align with left-leaning parties, they act out of vested interests rather than on strictly ideological grounds, working "to protect teachers' jobs, increase their wages and benefits, expand their rights, and restrict managerial discretion" (Moe, 2017, p. 28). In other words, unions are political actors that focus on these core issues for pragmatic reasons – because they directly matter in the lives of their members. Theoretically, this pragmatism should give them a degree of flexibility in terms of finding allies across the political spectrum who support public education.

Teachers' collective interests may or may not be aligned with students' interests. Moe's body of work suggests a belief that teachers' unions are often at cross purposes with students' best interests. If we accept that unions are reflexively opposed to reform and change, it remains up for debate and empirical study whether these reforms are actually beneficial for students. Education reformers tend to present their policy changes as implicitly positive and effective at raising student outcomes without waiting for long-term evidence or a full understanding of organizational change (Fullan, 2012). Unions are depicted, in contrast, as obstructionist because of their interest in protecting the status quo (Moe, 2015; Moe & Wiborg, 2017). Popular media infrequently highlights union concerns about the effectiveness of proposed reforms for students, though the media climate is now undergoing dramatic change and becoming more pro-union (Goldstein, 2019).

Furthermore, accountability reforms may have made the teaching profession less attractive (see: Ingersoll & Collins, 2018). Reforms may worsen the teaching environment, affecting teacher recruitment and retention, and negatively impacting student outcomes (as in Kraft, Marinell, & Yee, 2016). Or, they may improve teacher quality, if quality is defined as causing higher student achievement (Papay & Hannon, 2018). Still, teacher control and autonomy has been significantly associated with higher student proficiency for decades (Ingersoll, 2003; Ingersoll & Collins, 2018) and, more recently, teachers' evaluation of their schools. Considering the conflicting evidence available, it seems unwise to take autonomy away from teachers for the sake of raising student achievement without careful consideration of unintended consequences for the health of school communities in the long term, as well as the emerging racial match literature (e.g., Yarnell & Bohrnstedt, 2017). Students tend to do better with teachers who look like them. Another caveat is that quantifiable outcomes of interest may widely, making it difficult to generalize a dichotomy between teacher and student interests.

In the economics literature, Moe's assumption that what's good for teachers is bad for students may be more contested thanks to a recent working paper from Brunner, Hyman, and Ju (2019). The authors find that states with strong teachers' unions allocated resources towards raising teacher pay, which in turn raised student achievement in the aftermath of state finance reforms. They find a causal relationship between higher pay and higher achievement by treating changes in school funding policies as exogeneous shocks. In contrast, states with weaker unions used this increased funding from the state to alleviate local property taxes and boost teacher recruitment, instead of investing in pay

for the existing workforce (Brunner, Hyman, & Ju, 2019). This strategy did not boost achievement. This macroeconomic trend parallels what happened to the teaching workforce in Lawrence – turnover increased, and average years of experience decreased, though the quality of incoming teachers did improve (Papay & Hannon, 2018). But effectiveness increases throughout a teacher's career (Papay & Kraft, 2016). Consequently, the vested interest of teachers may not be all that different from the vested interests of students and families. More experienced, better compensated, and highly autonomous teachers may also improve student outcomes.

Next, I consider Moe's assertation and the available evidence that unions are reflexively opposed to increased accountability. I also present the evidence in favor of and against holding teachers accountable for student achievement, either through state law or individual district CBAs.

Union Opposition to Accountability and the Existing Evidence

When looking nationally, states with stronger unions and Democratic legislatures and governors clearly prefer fewer mandatory teacher observations. Democratic states also are less likely to rely on test scores to make hiring and firing decisions (Finger, 2017). But despite the greater popularity of accountability in less-unionized and more Republican-controlled states, the dawn and institutionalization (see DiMaggio & Powell, 1983) of evaluation policies that hold teachers accountable for student achievement measures remains politically controversial (Cheng et al., 2018; McGuinn 2012; White, 2018). This institutionalization reached its peak in the years following the 2009 announcement of RttT (Bleiberg & Harbatkin, 2018). RttT effectively incentivized states

to adopt the federal government's policy preferences for teacher evaluations and moved the boundaries of federalism (Halgren, James-Burdumy, & Perez-Johnson, 2014; McGuinn, 2012; Viteritti, 2012). RttT built upon the requirements of 2002's No Child Left Behind (NCLB), "the first major national accountability structure that mandated states to hold school and districts responsible for student achievement" (Wrabel et al., 2018, p. 118; Sunderman & Kim, 2007). Only two-thirds of U.S. states had substantive accountability systems prior to NCLB (Dee & Jacob, 2011; Ladd, 2017). The successor to NCLB, the Every Student Succeeds Act (ESSA), has enabled states to develop more nuanced accountability systems. Some states have retained their NCLB-era systems (Achieve, 2018), but more recent evidence suggests that gains have slowed (Alexander, Jane, & Kankane, 2017). Considering the evidence that stronger accountability can raise achievement for all even if it fails to close gaps based on race, ethnicity, and income (Polikoff, Korn, & McFall, 2018), achievement testing is still popular with the majority of the public according to the limited polling available (Cheng et al., 2018).

Despite this public support for some form of achievement testing, teachers dislike standardized testing, particularly when it is used for accountability purposes (Cheng et al., 2018; Ingersoll & Collins, 2018; Kaufman, Opfer, Bongard, & Kane, 2018). But there may be an ideological and partisan dimension to teacher accountability preferences that Moe minimizes in favor of his vested interest theory. In his framework, any managerial oversight of teachers is negative. Teachers, like most employees, seek to avoid any evaluative gaze (as in Foucault & Sheridan, 1975).

Non-teachers, however, divide along partisan and ideological lines. Two survey experiments suggest there is a great deal of nuance in the type of oversight – either observation-based or test-based – that Republicans or Democrats prefer. White (2018), in her survey experiment of state education policymakers, finds that knowing teacher preferences (for more or less test-based accountability) pushes Republican respondents more strongly against the teachers' preference. Republican and conservative disapproval of teachers' unions is most likely a cause (Cheng et al., 2018; Moe & Wiborg, 2017). Meanwhile, there is growing evidence that individual Democrats prefer the position that the local teachers' union holds, and that voters are increasingly aware of these default partisan positions (Finger, 2017; Moe, 2017; White, 2018). Teachers may similarly be divided along ideological and partisan lines, or they may have policy preferences based on their own abilities to meet accountability targets.

Individual teachers, however, are sublimated into their local union. Moe is correct in noting that unions as a whole typically advocate for the least amount of oversight possible – whether that be fewer observations or fewer tests. An outcomes-based system of accountability is inherently less expensive and, in the minds of conservatives, more objective (Chambers, Schlenker, & Collisson, 2013; Chubb & Moe, 1991; Tetlock et al., 2013). A process-based system using in-class observations, in contrast, provides useful feedback without reducing autonomy and trust among effective teachers (Firestone, 2014). But observations take more time and cost more money. Unions prefer restrictions against both types of accountability (Moe, 2015).

In short, liberals believe that evaluating teachers based solely on test scores is unfair and in keeping with the teachers' union position. Liberals emphasize the legacy of incoherence in public schooling and a lack of resources to help teachers teach standards-aligned curriculum (if there is any curriculum at all) (Cohen, Spillane, & Peurach, 2018). Liberals worry that teachers will teach to the test and engage in harmful gaming-the-system behaviors (e.g., Booher-Jennings, 2005). Conservatives take the opposite tack, believing public sector unions to be wasteful and inefficient. Conservatives believe that unions protect teachers who would be easily be fired in the private sector or in non-unionized environments. They believe that teachers' feet should be "held to the fire" (Tetlock et al., 2013, p. 23). In developing hypotheses for what might predict the presence of certain CBA items in Pennsylvania, I rely heavily on these generalizations.

Unions, however, have been more receptive to accountability when funding is attached to reforms, as opposed to increased pressure without increased support (Ladd, 2017; McGuinn, 2012). Irrespective of the type of accountability, public schools in the United States cannot receive increased funding without increased oversight (Mehta, 2013), even if one believes that accountability systems trend towards dysfunction or, as Au (2016) argues, reinforce rather than remediate racial inequality. Public accountability for government-paid employees within a democratic system is a given. As Cohen, Spillane, and Peurach state, "School systems are open systems and depend on their environments…they operate in a complex and pluralistic institutional environment that contains different and often divergent pressures for action" (2018, p. 209). Similarly, Tetlock and colleagues (2013) write, "when organizations make flawed decisions, one

rarely needs to wait long for those harmed to demand 'more accountability'" (p. 3). In the case of schooling, legislators and parents are the ones often making these demands. The question that remains is how to meet this public demand during a period where there has been a "dramatic loss of public confidence" in public schools (Cohen, Spillane, & Peurach, 2018, p. 204). Unions that resist accountability pressures emerge as a scapegoat for the perceived shortcomings of public education. They are a convenient rhetorical tool, a means by which to avoid blaming individual teachers. Only recently are we beginning to see a turn in public opinion in favor of teachers' unions during an era of increasingly successful teacher strikes (Goldstein, 2019).

I turn next to the rationales behind accountability and the empirical evidence. By emphasizing accountability, reformers may be attempting to do on the back end what they failed to do on the front end (Mehta, 2013; p. 7) – focusing on outputs (student achievement) rather than inputs (increased funding and resources). Perrillo (2004), in her history of the professionalization of the teaching force, puts accountability in a broader historical context as far back as the dawn of the 20th century: "The effort to transform the teacher, then, ran parallel to the larger problems they could not change within the schools (p. 350)." Other historians argue that accountability metrics pre-date even the founding of the United States (see Gamson, 2015). But this context does not fully appreciate the furor surrounding today's critics of test-based accountability, and why unions might be opposed for reasons outside of self-preservation. Unions might also be opposed because teachers may know accountability is bad for students despite marginal achievement gains (see Dee & Jacob, 2011; Springer, 2008 for evidence of these gains).

The notion that accountability is best for students (see Hanushek, 1986 for one of the first such arguments) has come under increasing fire from multiple studies across a diverse range of methodologies. Ladd (2017) considers accountability policies such as No Child Left Behind to be deeply flawed, as they focus too heavily on gains in short-term academic outcomes (found in studies such as Dee & Jacob, 2011; Dee, Jacob, & Schwartz, 2013; Hanushek & Raymond, 2005). Using even stronger language, Au argues that "high-stakes testing cannot dismantle racial inequality because it fundamentally and materially advances the project of increasing racialized injustice" (2016, p. 41). Test-based accountability allows "education to be reconstructed around a simple model of commodity production and consumption" (p. 42). Much qualitative research has also documented the harmful practices that occur when unions do give in to increased accountability (e.g. Booher-Jennings, 2005; see Steinberg & Quinn, 2015). Viewed through these lenses, the resistance of teachers to accountability-based reform may represent firsthand knowledge of negative student experiences with testing rather than mere self-interest.

Teachers and teachers' unions are also not alone in their opposition to test-based accountability. Administrators may also play a crucial role in resisting. District leaders are continually reconstructing state education policy (Woulfin, Donaldson & Gonzalez, 2016; Daly & Finnigan, 2011) while they accommodate the institutional memories of their teaching force (Goodson, Moore & Hargreaves, 2006) and navigate emotional networks (Fullan, 2016). Principals have been found to either enhance or subvert accountability systems to better meet institutional needs (Donaldson & Woulfin, 2018).

They may work on teachers' behalf to shield teachers from outside influence and political pressures (Weick, 1974).

Considering the multiple stakeholders who might oppose, subvert, or enhance accountability, it seems unlikely that the accountability provisions in CBAs will demonstrate significant relationships to student achievement. Much of the accountability policy in Pennsylvania is already set at the state level, as I discuss in the next subsection (PDE, 2018). While the econometric literature is somewhat positive, the qualitative literature on the topic suggests severe unintended consequences (Edgerton, 2019a). I accept the notion that unions prefer less accountability, in all its forms, but problematize the notion that accountability is universally good for all students. Furthermore, there may be no relationships between accountability found within CBAs and student achievement, as there are many, many steps between the document itself and actual implementation (Spillane, 2009). And the econometric evidence, even when implementation is ideal, is decidedly mixed (Ladd, 2017). Next, I turn to the specifics of the Pennsylvania policy context, the commonwealth's long history of teacher strikes, and the specifics of the collective bargaining process.

Collective Bargaining in Pennsylvania

The legal foundation of public education in Pennsylvania is the Public School Code of 1949, which the state legislature continually amends. Teachers' unions first formed in Pennsylvania with the passage with the passage of Act 45 and the Public Employee Relations Act in 1970, which provided teachers with the ability to strike. For decades, Pennsylvania teachers exercised this right more than teachers in any other state.

This activism generated a set of policy responses that make the study of this state a unique contribution to the collective bargaining literature. From 1972 until 1992, there were an average of 27.6 strikes annually; this number dropped to 8.6 strikes during the period of 1992 to 2007 (Zwerling, 2008). There have been 131 strikes from 1999 to 2018, and Pennsylvania is one of only 12 states that explicitly legalizes them (Brandt, 2018). Rather than attempt to ban strikes altogether, state law has instead moved to limit the duration of strikes by requiring teachers to make up any days lost in order to meet the 180-day instructional minimum (Zwerling, 2008). According to Act 88, which was passed in 1992 to curtail the prevalence and duration of teacher strikes, unions must now give 48 hours' notice prior to a strike. Despite these reforms, Pennsylvania still leads the nation in the sheer number of district-level teacher strikes, though it is also among the Top 10 in the country for the sheer number of school districts, which makes this metric somewhat misleading (Brandt, 2018).

Existing agreements prohibit strikes during the duration of the agreement; strikes thus occur when the two parties (the union and the school board) cannot agree to a new CBA. In Pennsylvania, two parties are signatories on the final CBA – the union president, and the school board president. State law has purposefully defined the negotiation timeline to maximize cooperation between these two groups. Negotiation must begin by January 11th of the last year of the contract, and mediation must occur on or before February 25th of the same year. Both parties must accept or reject fact-finding reports delivered by the Labor Relations Board, when requested (Pennsylvania State Education Association, 2018). This strict timeline aims to lessen the chance that teachers will strike.

Thus, while Pennsylvania has prescriptive mandates as to how teachers must be evaluated, they retain robust collective bargaining rights. Strikes may thus result from fiscal challenges and the need to make hard trade-offs, particularly in underfunded areas of the state. The state has notoriously inequitable school funding, which may account for the prevalence of district-level strikes in the state (Steinberg & Quinn, 2015). A strike resulting from a failure to negotiate a new contract may also stem, I argue, from sharp partisan differences, if a heavily Republican district finds itself doing battle with a Democratic teachers' union. With these structural realities in mind, I turn to the items in Pennsylvania CBAs that might demonstrate relationships to student achievement, before discussing the importance of partisanship and my theoretical framework.

Collective Bargaining Agreement Items and Student Achievement

The importance of teachers in boosting student achievement has been long established (Hanushek, 2011; Rockoff, 2004) in addition to being self-evident to those who have taught. Throughout this dissertation, I test the alignment between teachers vested interests and student interests in Pennsylvania by using student achievement as a proxy for students' best interests. This is hardly a perfect proxy considering the racialized history of achievement testing (Kendi, 2017), but it is one of the few outcomes available for quantitative analysis. Using achievement as a starting point, I consider whether student and teachers interests might be more in sync than is popularly believed. I also consider whether some CBA items are better suited than others for improving student academic outcomes. I hypothesize that most items will demonstrate no significant

relationships considering the many implementation steps and intervening factors between CBAs and student scores.

Collective bargaining research on student outcomes remains relatively underdeveloped and often focuses on large national datasets and policy changes rather than on the substance of district CBAs. For example, a recent study by Lovenheim & Willén (2019) finds negative long-term outcomes, including employment and earnings, for Black and Hispanic students in states that mandate collective bargaining. Studies such as these rely on exogenous shocks to leverage econometric techniques, but they do not examine the intricacies of CBA items.

A few scholars have pioneered codifying the actual content of CBAs in four states – California, Washington, Michigan, and Ohio – or in large urban districts only (Goldhaber, Lavery, Theobald, D'Entremont, & Fang, 2013; Hess & Loup, 2009; Marianno et al., 2017; Moe, 2009; Strunk et al, 2018; Strunk & Grissom, 2010; Strunk & Reardon, 2010). These studies measure the extent to which the CBA specifies or restricts certain activities, such as teacher evaluations. Recently, a study of a new teacher evaluation systems by Stecher and colleagues (2018) concluded that these changes to CBAs in three large school districts (including Pittsburgh) had null and, in some cases, even negative effects on achievement. Evidence as to the influence of CBAs on student achievement remains a mixed bag and differs based on the methodology and location of the study. Marianno & Strunk (2018) find that in California "the relationship between CBA strength and student achievement is persistently negative and small, or null, but

never significantly positive" (p. 93). None of these studies, however, consider Pennsylvania.

In Pennsylvania, CBA items vary on core issues despite strong state accountability mandates. I focus on these items in my data collection, as they represent the substantive variations in the CBAs that could demonstrate relationships to student achievement. In terms of accountability and seniority, CBAs vary in the negative consequences ascribed to low teacher evaluation ratings. More recent amendments to the School Code have increasingly specified how these evaluation ratings are to be used. In 2018, the legislature passed Act 39, which de-emphasized teacher seniority as grounds for furlough during economic hardships. Districts are now able to dismiss teachers with the lowest ratings rather than rely on seniority exclusively; last-in, first-out is no longer the law of the land. Previously, seniority protected teachers from layoff regardless of their individual evaluations. This recent amendment to the School Code is beyond the scope of the most recent available data, but it is an important one to watch in the coming years, even if in reality few teachers actually receive Failing or Needs Improvement ratings (Steinberg & Kraft, 2017).

While there is much legislative activity around seniority rights when districts are in economic distress, Pennsylvania districts have more discretion when it comes to initial hiring. A few districts (17%) engage in seniority-based hiring, where existing bargaining unit members have preference for new positions in order of years of experience in the district. Other districts (24%) have a bit more flexibility but are still required to consider all existing teachers who apply for current vacancies. They must provide in writing an

explanation should they chose new applicants over existing teachers. The remaining majority of districts (59%) allow principals and superintendents to make hiring decisions as they see fit without consideration for seniority or current union members.

Critics point to these seniority rules as stifling innovation within the education sector (Moe, 2017). Institutional memory is undervalued in education reform circles (Goodson, Moore, & Hargreaves, 2006), but de-emphasizing the importance of teacher experience also contradicts the teacher effectiveness literature, in particular the finding that effectiveness may not be portable across district lines (Papay & Kraft, 2016). The removal of seniority rights is meant to address what is considered to be a teacher quality gap, but Koski & Horng (2007) find "no persuasive evidence that the seniority preference rules independently affect the distribution of teachers among schools or exacerbate the negative relationship between higher minority schools and uncredentialed and low-experience teachers" (p. 262). However, Anzia & Moe (2014) find more recent evidence that contradicts this earlier study.

The matching of teachers to open positions is a two-sided process. When given the flexibility, districts are more likely to hire teachers with higher measures of quality across a range of characteristics, such as training, college selectivity, years of experience, and prior demonstrated effectiveness. Teachers are also likelier to choose districts with the best working conditions (Boyd, Lankford, Loeb, Ronfelt, & Wyckoff, 2011). Thus, there is a great deal of self-selection – more effective teachers stay in their jobs, and more effective schools are better at hiring more effective teachers and retaining them. Seniority

rules may play a limited role in this complex process, making potential relationships to student outcomes weak or nonexistent.

Class Size

Pennsylvania districts also have discretion as to whether and where to set maximum class sizes. There is only one hard cap for pre-K (20 students) and no statewide caps in K-12 (Zinth, 2009). Research on this topic is mixed according to Chingos and Whitehurst (2011), as the pool of credible studies is relatively small. Older meta-analyses suggest null effects for this expensive policy mandate (Hanushek, 1997). However, positive effects on student achievement and class size reduction have been found in Tennessee (Krueger, 2003), Texas (Rivkin, Hanushek, & Kain, 2005), and California (Gilraine, Macartney, & McMillan, 2018), while other studies have found negative or null effects in Connecticut (Hoxby, 2000a) and Florida (Chingos, 2012). Class size reductions may be most effective for low-income and minoritized students (Mathis, 2017). However, they may be counterproductive for others, and a more recent meta-analysis finds null effects of class size reductions (Filges, Sonne-Schmidt, & Nielson, 2018).

Here it is important to note that I am studying relationships to maximums, not the effects of a new class size reduction policy. Only 25 urban and suburban districts in Pennsylvania specify a class size maximum. Consequently, I anticipate few relationships between class size and student achievement when looking across Pennsylvania considering its many small rural districts. Class size reduction is typically a concern for large urban school districts, which I do not study in isolation.

Teacher Prep Periods

In addition to negotiating class size, districts individually negotiate with local unions for how much preparation time teachers are guaranteed every day. Of course, what happens during preparation time (which may or may not include common planning time, or professional learning communities [PLCs]) matters a great deal (Desimone & Pak, 2017; Desimone, 2009; Dever & Lash, 2013). It is impossible to measure how teachers might be allocating their time through this particular study design. Collaboration during a shared planning period, however, still has been shown to predict student achievement (Reeves, Pun, & Chung, 2017). I argue that the existence of planning time may be a proxy for collaboration – a necessary structure for it to occur. Most teachers have around 45 minutes of guaranteed planning time nationally per day, or about 225 minutes per week (Merritt, 2017; NCTQ, 2012). I hypothesize that the absence of guaranteed prep time will increase barriers to teacher collaboration.

However, there is not a great deal of variation across districts for this measure since most districts have similar prep time policies. According to my coding of the CBAs, 75% of Pennsylvania district specify prep time in grades 4 and 8. Those smaller rural districts with short contracts that do not specify planning time most likely still provide a prep period to their teachers, even if they are not contractually obligated to do so. There is notably one outlier district that guarantees nearly double the national and stage average of elementary school prep time (the average is 221 minutes for Pennsylvania, excluding those 122 districts that do not specify the time). Baldwin-Whitehall, located outside of Pittsburgh, guarantees elementary grade teachers 525 minutes of preparation time in a

week, which is exceedingly rare. High school teachers in Baldwin-Whitehall, in contrast, receive a more typical 200 weekly minutes (one class period of 40 minutes). There is a weak but significant positive relationship (Pearson's *r=0.09)* at the 0.05 level between prep time and math and ELA achievement in grades 3-8. Considering the skewedness of this measure, I do not expect the planning time variable to be significantly predictive in my models.

Teacher Salary, Structure, and Bonuses

Next, I consider items that have occupied much more political attention and empirical study – teacher salary structures. Higher teacher salaries have been linked to better outcomes for students globally (Dolton & Marcenaro-Gutierrez, 2011; Figlio & Kenny, 2007), as has performance pay when looking at large datasets (Woessmann, 2011). More localized experiments in teacher performance pay have had mixed results, leading to null, small negative, or small positive effects depending upon their location and methodology (Chiang, Wellington, Hallgren, Speroni, Herrmann, Glazerman, & Constantine, 2015; Dee & Wyckoff, 2015; Podgursky & Springer, 2007; Shifrer, Turley, & Heard, 2017; Springer, Ballou, Hamilton, Le, Lockwood, McCaffrey, Pepper, & Stecher, 2011; Stecher et al., 2018; Steinberg & Sartain, 2015). The most recent study on this topic, in Tennessee, finds that bonuses of $5,000 can boost student achievement (Swain, Rodriguez, & Springer, 2019).

The level and structure of salaries may play a key role in the attraction, retention, and motivation of teacher. Conversely, the intrinsic motivation of teachers to enter the profession (see Firestone, 2014) may be weakening the impact of performance-based pay

by conflicting with collegiality and how individuals understand the purpose of schooling (Ingersoll & Collins, 2018; Mintrop, Ordenes, Coghlan, Pryor, & Madero, 2018). The existing empirical evidence suggests that performance-based pay schemes may cause less effective teachers to leave the profession. They may also increase the quality (as defined by test scores) of incoming teachers (Adnot, Dee, Katz, & Wyckoff, 2017; Papay & Hannon, 2018). In contrast, the impact of teacher pay raises without performance incentives attached is far less studied in the recent literature. In the United States, there is an implicit assumption that with increased pay must come increased accountability (Mehta, 2013).

One has to look outside of the United States for experiments with unconditional pay raises. Developing nations are much more likely to implement across-the-board teacher increases rather than attempt to reform the existing salary scale and incentives. In Indonesia, a raise “significantly improved teachers' satisfaction with their income, reduced the incidence of teachers holding outside jobs, and reduced self-reported financial stress” (dee Ree, Muralidharan, Pradhan, & Rogers, 2018, p. 993). However, these improvements did not translate into student achievement effects, which is consistent with the literature in other countries with similar pay experiments that find no relationship between teacher pay and effectiveness (Bau & Das, 2017; Hanushek, 1986; Muralidharan & Sundararaman, 2011). While I cannot study the effects of a pay increase with this study design, I can examine those districts in the top quartile of the state.

Returning to the United States, unions across sectors have had a significant equalizing effect on income in this country since 1936 (Farber, Herbst, Kuziemko, &

Naidu, 2018). A recent meta-analysis of unions and pay increases suggests that teachers' unions are marginally effective at increasing wages by about 3%. However, unions were more effective at securing large wage increases during earlier periods of reform in the 1970's (Merkle & Phillips, 2018). Prior work shows that unions are effective at securing large increases for veteran teachers but not for newer teachers (Winter, 2011). Considering recent teacher strikes in low-paying states (Campbell, 2019), we may be at the dawn of a new era of union effectiveness. Despite complaints that generous union benefits siphon off critical resources from under-resourced schools, American teachers remain underpaid when looking at comparable countries. To match the teacher salaries of even a poorer country such as Poland, the United States would have to implement an average salary increase of 20% increase (Startz, 2016).

Considering these structural and political barriers, much research instead focuses on changes to pay structures rather than average pay. Research on the effects of bonuses based on additional degrees, credits, National Board Certification (NBC), and longevity, is mixed. Some Pennsylvania districts offer small ($500) longevity bonuses for every ten, twenty, thirty, and even forty years in the district. There is no literature on longevity bonuses, which may be a holdover from decades-old contract negotiations. These incentives occur in spite of evidence that teacher effectiveness plateaus and may even decline in later years (Chingos & Peterson, 2011), though Papay and Kraft (2016) establish that this plateau effect masks the incremental gains made in later years. Effectiveness among veteran teachers increases by 35% from year 10 to year 30 (Papay & Kraft, 2016).

Aside from years of experience, NBC is a stronger signal of effective teaching, even if this effectiveness does not stem from undergoing the certification process itself (Goldhaber & Anthony, 2007). Incentives for teachers to become National Board certified have been effective at increasing the number of certified teachers in Washington State, though this increase in NBC teachers did not lead to increases in student achievement (Cowen & Goldhaber, 2018). As with much of the literature on bonuses, only a handful of states with excellent data collection practices have been studied, and Pennsylvania is not represented.

Finally, Pennsylvania districts may also choose, through a salary scale, to place a higher or lower amount of emphasis on graduate degrees and credits. The number of columns in the salary schedule, which I call "narrowness", reveals the extent to which teachers are awarded pay increases for graduate credentials (as opposed to higher evaluation scores, for example). The shortest number of columns that a district may have is two – one for Bachelor's, and another for Master's degree holders. A narrow scale removes the incentives for teachers to earn additional degrees and credits. Districts with narrow scales, such as Pittsburgh, place a greater emphasis on evaluation scores instead of credentials. Pittsburgh has only two columns and ten rows, which allows teachers to quickly advance up the scale (Stecher et al., 2018). To proponents of evaluation reform, this narrow scale and rapid advancement is ideal for recruiting and retaining the most effective teachers.

But this approach is still rare in Pennsylvania and has not diffused from urban centers. Many districts offer increases for every 15 credits earned, while a few offer

salary increases for as few as 6 credits earned. This differentiation may incentivize teachers to rack up graduate credits and professional development points (PDPs) despite the lack of evidence for their importance in the literature. The lack of a relationship between teacher effectiveness and graduate credits is well-established (Chingos & Peterson, 2011).

Whether these incentives to increase teacher credentials are effective at doing so remains up for debate. Relevant studies examine the effects of teachers' credentials, but they do not study the effectiveness of district incentives that aim to encourage teachers to obtain them. Clotfelter, Ladd, and Vigdor (2009) find that having licensed and highly qualified teachers positively affects student achievement in North Carolina high schools. Kane, Rockoff and Staiger (2008) echo these findings in New York City, as do Darling-Hammond, Holtzman, Gatlin, and Heilig (2005) in Houston, Texas. However, initial certification and licensures are not equivalent to additional credits earned while teaching, and districts are increasingly less likely to pay for these professional development opportunities. None of these studies have used data from Pennsylvania.

I next discuss the literature on school boards and partisanship in Pennsylvania before concluding with my synthesis of this literature and the theoretical framework guiding this research.

School Boards and Partisanship in Pennsylvania

When negotiating the provisions discussed in the previous section, teachers' unions have major advantages at the bargaining table compared to local governments. School board members are portrayed as inept negotiators against powerful teachers'

unions (Kirst & Wirt, 2009). Unions, conversely, shape policy from the top-down through extensive lobbying of state and national governments and, in some ways, set the terms of the debate (Moe, 2017). Local school boards are more constrained in their influence and much smaller in number. Structurally, the U.S. is unique among nations in not bargaining for teacher contracts at a higher level of government than the school district (Moe, 2017). Leaving bargaining up to individual districts puts local governments at a distinct disadvantage.

In Pennsylvania, school board members – called school directors – are not able to coordinate across district lines, which theoretically could help with teacher retention. There are nine directors for every Pennsylvania school district. They are elected every two years in odd years, and they alternate between having either 4 or 5 members up for re-election. Off-year elections are notoriously low in voter turnout (Fiorina, 1978). Educators themselves may also be elected to school boards and further increase the union's political advantage. In a study of California school districts, educators on school boards were able to increase teacher wages and reduce charter school enrollment, but they did not affect student achievement either positively or negatively (Shi & Singleton, 2018).

School directors may or may not be operating from a distinctly partisan lens, and voters may be less aware of their policy preferences. Recent research suggests that Democratic school board members pass policies that decrease school segregation (Macartney & Singleton, 2018), while ideologically mixed boards exhibit greater interpersonal conflict (Grissom, 2010, 2014). School board candidates at the top of the

ballot are more likely to be elected than candidates at the bottom of the ballot; school board elections are extremely low information environments (Meredith & Grissom, 2010; Shi & Singleton, 2018).

Currently, Pennsylvania school board candidates can cross-file as both a Republican and as a Democrat when running for office, which can confuse voters as to the candidate's actual political party affiliation. But candidates that win the nomination of both parties receive a large boost of 14-19% of the vote share (Meredith & Grissom, 2010). Candidates who are not cross listed as both Democrat and Republican rarely get elected and are virtually always ranked last (out of 4 or 5 possible choices). Most candidates run unopposed (Meredith & Grissom, 2010) and tend to come from more affluent backgrounds, where they might use the office as a steppingstone to a higher office (Bartanen, Grissom, Joshi, & Meredith, 2018; Kirst & Wirt, 2009). A recent bill aims to end the practice of cross-filing and thus increase partisan signals to voters in these elections, but it has not yet passed (Esack, 2018).

There are 168 rural, 75 town, 236 suburban, and 20 urban districts in Pennsylvania, according to NCES (2018). In a state such as Pennsylvania with more than 100 small rural districts, party cues are considerable in these low-information school director elections according to regression discontinuity estimates (Meredith & Grissom, 2010). Directors may be less responsive to constituencies in low turnout elections. The relationship between directors and voters is bidirectional but unbalanced. Directors may send crucial partisan cues to voters in order to frame complex contractual debates (Campbell, Converse, Miller, & Stokes, 1980; Chong & Druckman, 2007).

Though partisanship has been part of the United States since its founding, partisan conflict may be expanding into previously bipartisan areas of education policy. Nationally, education policy issues are increasingly polarized along party lines according to state-level studies of union and partisan influence (e.g., Finger, 2017) and opinion polling (Cheng, Henderson, Peterson, & West, 2018). And there is evidence to support the nationalization of education party opinion irrespective of local policy contexts (Edgerton & Desimone, 2018; Supovitz, 2017). Voters now better understand the default Democratic position (pro-union, against increased teacher accountability) as well as the default Republican position (anti-union, in favor of more teacher accountability) as partisan elites sort themselves on either side of these issues (Levendusky, 2009; Finger, 2017; White, 2018). I expect that areas with more Republican voters and school board members will prefer lower teacher wages and higher accountability (Brunner, Hyman, & Ju, 2019; Tetlock et al., 2013). Even coronavirus efforts are evaluated through a partisan lens (Gadarian, Goodman, & Pepinsky, 2020).

Considering these partisan political factors, I integrate the political science literature into the existing econometric literature within the theoretical framework described below.

Theoretical Framework

Ultimately, the question of whether teachers' unions are "good" or "bad" for students is a normative one without overwhelmingly strong empirical evidence. On the positive side, teachers' unions may simultaneously protect student interests, increase professionalism, improve teacher quality (Han, 2016), and check administrative and

bureaucratic power (Casey, 2006; Ingle & Wisman, 2018; Ravitch, 2006-2007). On the negative side, unions may focus too much on individual teachers, undermine accountability, and preference seniority over effectiveness (Ballou 2000a, 2000b; Hill, 2006; Ingle & Wisman, 2018; Moe, 2015). I theorize that these positive and negative relationships appear among individual contract items, representing both vested interests as well as more direct impacts on student outcomes, such as class size.

Teacher qualification indicators – pertaining to bonuses, professional growth plans, preparation time guarantees, hiring, and dismissal – should in theory improve teachers' working conditions, which may increase teachers' satisfaction with jobs and, indirectly, improve their instruction. But classroom indicators related to class size and teacher evaluation should more directly influence student outcomes. Using these two constructs (classroom v. teacher qualification) as well as the overall strength of the contract (classroom + teacher qualification), I complicate the notion of teachers' union vested interests and seek to find where they overlap with that of students and families. It may be the case that there is more overlap than is commonly believed.

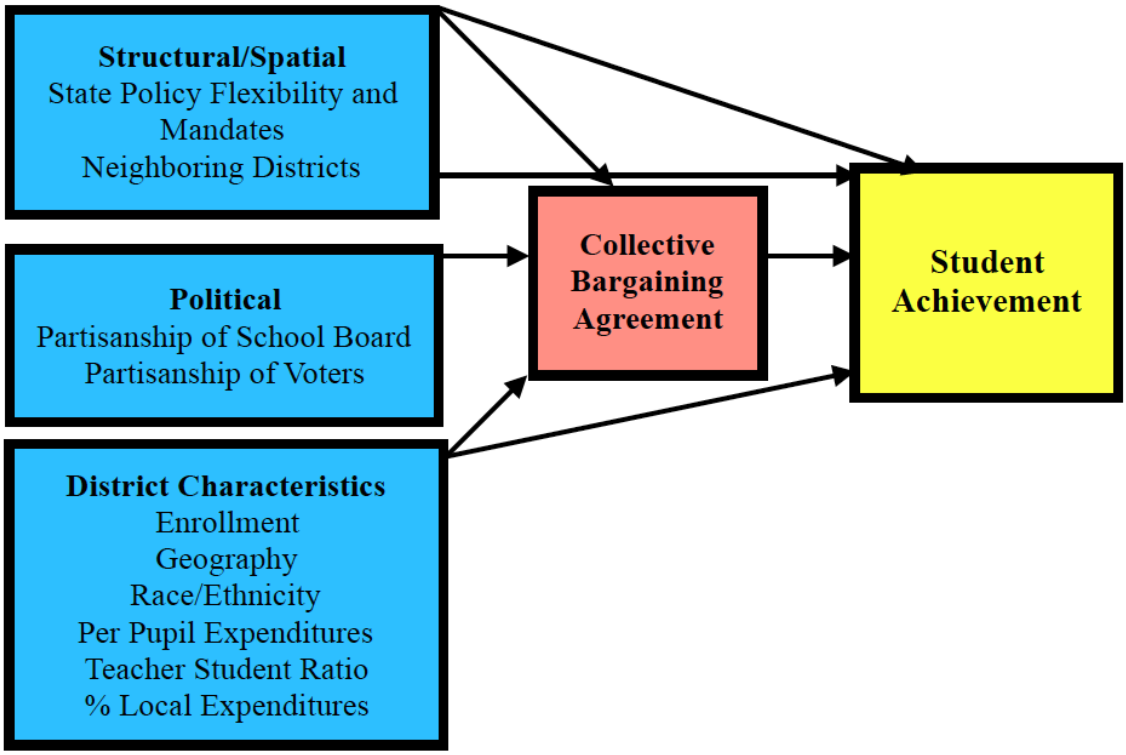
Though a major limitation of my study is in examining only a single year of data, these contracts are remarkably stable over time based both my reading of contracts that changed in 2016-2017 and the existing literature. CBAs are often copied-and-pasted from year-to-year, with the exception of inflation-adjusted salary schedules and benefits (Ingle & Wisman, 2018).

Referring back to Table 1 and the literature discussed in the previous sections, I hypothesize that negative consequences for poor evaluation scores and National Board

certification bonuses will relate to student achievement and graduation rates. Cumulatively, classroom items may also demonstrate positive relationships to these outcomes. I consider the remaining items centered on issues of seniority, graduate school credentialism, and preparation time to be less directly related to student academic outcomes. These items may have null or negative estimates on achievement, which would be consistent with the existing literature on these topics. However, there are many steps between teacher working conditions and student achievement. Thus, I anticipate that significant relationships will be small in magnitude.

Additionally, the existing CBA studies (Strunk et al., 2018) do not consider the influence of political partisanship, given that collecting information on school board elections is time-intensive and elections are typically nonpartisan. I incorporate these sociopolitical factors into this dissertation by looking at both registered voters and school boards. I have discussed the structural constraints of Pennsylvania's policy environment in depth and highlight that these state-level structures set both a ceiling and a floor on much of what can appear in CBAs. Governance arrangements determine union involvement in deciding upon CBA provisions (Finger, 2017). I summarize all of these variables in Figure 3 below.

Figure 3. Theoretical Framework for Understanding Relationships Between District Characteristics, CBAs, and Student Achievement



The structural/spatial component of the framework includes not only the state policy context as outlined previously but also the awareness of districts to their neighboring competitors. In a state like Pennsylvania, districts are competing for access to highly qualified teachers, particularly in hard-to-staff areas, and competing within a grossly inequitable property tax system that is among the most imbalanced in the nation (Candelaria & Shores, 2019; Steinberg & Quinn, 2015). The school choice literature is also rife with examples of how districts change their behaviors to maximize efficiency when competing for students (e.g., Hoxby, 2000b; Rincke, 2006). In terms of teacher labor markets, Jackson (2012) finds that the opening of new charter schools within a district decreases hiring in difficult-to-staff schools and reduces teacher quality in district

schools. Similarly, Feng, Filglio, and Sass (2018) show that teachers leave schools that receive failing grades. Thus, it stands to reason teachers would also leave districts that offer subpar contracts. While they may not have detailed knowledge if the intricacies of the contract, they are no doubt aware of what they might earn in a neighboring district.

A spatial perspective not only addresses the reality of how information crosses district lines but also illuminates the implications of Pennsylvania's gerrymandered and segregated districts. In *Seeking Spatial Justice*, Edward Soja (2013) writes that "an assertive and explanatory spatial perspective helps us make better theoretical and practical sense of how social justice is created, maintained, and brought into question as a target for democratic social action…whatever your interests may be, they can be significantly advanced by adopting a critical spatial perspective (p. 2). Though I do not often center social justice in this work, I emphasize that the wealthiest districts can pay the largest salaries and, consequently, deprive historically disadvantaged students of those teachers who can command the highest salaries. These are teachers who are in the highest demand, either because of their specialized nature or because of their prior effectiveness.

The reality of school funding in Pennsylvania constitutes a structural spatial injustice, and this study can be viewed as an attempt to place that injustice into its proper context. Prior studies have demonstrated the reality of wage spillover effects (e.g., Babock, Engberg, & Greenbaum, 2005), and contractual inequities may further drive teacher equality inequities. If there are significant differences between districts in terms of contracts, these differences may in turn be reflected in student outcomes. And if

neighboring districts are having to match each other in terms of their contractual obligations, illuminating this dependency may lead to more equitable policy solutions – by proposing statewide salary caps, to give one possible example.

For the political component of the theoretical framework, I propose two means of measuring political factors by looking at school board members and voters. These may be highly correlated, or there may be significant gaps in representation when comparing the partisanship of the individuals at the top with the electorate, as previous work has found (Meredith & Grissom, 2010). Finally, I draw upon the literature in other states to include the same relevant variables and district characteristics that might influence the CBA negotiation process as well as student achievement (see Strunk et al., 2018). I hypothesize in keeping with Strunk and colleagues (2018) prior work that there may no relationships between overall measures of CBA strength and student achievement. However, there may be relationships when looking at individual items, which I have grouped into teacher qualification and classroom indicators. This grouping is unique in the existing literature. I describe the methods for ascertaining these relationships in the next chapter, methods I have chosen to reflect the priorities of the theoretical framework.

CHAPTER 3: METHODS

Pennsylvania's Right-to-Know laws have made obtaining every CBA relatively easy because of this legal mandate to provide them. Consequently, I was able to read and code every contract instead of using a sample, as has been the case in other studies (e.g., Strunk et al., 2018). The most recent contracts were provided in a Dropbox folder thanks to assistance from the Commonwealth Foundation; I requested additional contracts from those districts that negotiated new ones for the 2017-2018 school year.

Recent studies by Strunk and colleagues (2018) have made the work of quantitizing contracts (see Sandelowski, Voils, & Knafl (2009) for a discussion of quantitizing) more manageable by reducing the number of data points needed to construct a strength scale. Earlier work by Strunk and Reardon (2010) collected data on more than 600 contract items in California, which was made possible by extensive outside funding. My coding process did leverage additional funding to hire four graduate students, as there were 24,157 pages to read for the most recent available contracts. Fortunately, many of these pages concerned themselves with retirement and medical benefits, which are beyond the scope of the study.

I personally coded 66% of the contracts and trained four other coders, who were also students in my graduate-level summer course. I instructed them to grade three contracts, and then send their coding to me to check for accuracy. I repeated this process until they were coding consistently and accurately. I also graded at least 10% of each of their assigned contracts once completed to further check for interrater reliability. They, in turn, reviewed a random 5% of my previously coded contracts for accuracy. Many of the

most confusing items with lower interrater reliability (the grievance procedure questions) are not included in the final analysis or in Table 1 because of a lack of variation. Throughout my analytic process, I have continued to spot-check contracts to ensure accuracy.

After the completion of the first round of coding, I also examined those contracts that expired in 2017 – 107 additional contracts. I submitted additional Right-to-Know requests to each of these school districts in order to examine variation over time for this targeted sample (21%). I also re-examined the prior contracts for each of these districts, adding additional reliability.

As for measuring the overall strength of CBAs and their high-profile provisions, research by Strunk & Reardon (2010), Strunk & Grissom (2010), Goldhaber and colleagues (2013), Marianno and colleagues (2017), and Strunk and colleagues (2018) have advanced the study of CBAs by using Partial Independence Item Response (PIRR) to analyze contracts using a common "survey" of items (Reardon & Raudenbush, 2006). Essentially, they developed a checklist. These common items (collectively, "restrictiveness") provide a solid foundation for the coding of contracts in a manner that is valid, reliable, and replicable. For example, a CBA cannot specify class size without first mentioning class sizes in the contract. The PIRR technique treats contract provisions as gateways, where provisions build upon each other, resulting in large, restrictive contracts in urban districts compared to smaller, less specific contracts in rural districts. Geography alone, however, does not fully explain CBA variations in the previously studied states.

I pause here to note a limitation of my work, as I did not use the PIRR technique, which would have required many more coders and additional data collection. However, by limiting my data collection to high-profile items adapted from Strunk and colleagues (2018), I believe I have captured the items of the greatest relevance to policymakers and eliminated unnecessary redundancies. I consider my approach to be appropriate considering Pennsylvania's unique and already highly policy restrictive environment. Even among these high-profile items, I found remarkably little variation, as Chapter 4 will reveal.

Aside from my own generated CBA database, the Pennsylvania School Data Project, housed at Research for Action (RFA), consolidates the existing data sources from the Pennsylvania Department of Education (PDE), Civil Rights Data Collection (CRDC), and the Common Core of Data (CCD). I have combined all of the high-profile contract items using the database from RFA (2018), which provides all of the information on district demographics referenced in my theoretical framework and in the previous chapter.

For research question 1, I use free QGIS software to generate maps for each of the high-profile provisions as well as other key characteristics. I merged a publicly available shapefile with the results of my coding, using unique district identifiers (OpenDataPA, 2018).

Models

For research question 2, I use the regression model below to predict the presence of each item (or the continuous outcome). Table 1 lists all of the binary outcomes as well

as the continuous outcomes (rows, columns, average teacher salary, preparation time, and class size). I use the quasi-maximum likelihood estimator (QML), the generalized spatial two-stage least squares (GS2SLS), and the spregress command in Stata. I incorporate the spatial dependence of having a neighboring school district with similar contract items using spatial autoregression using the following equation:

$$y_c = a + \beta_i X_i' + \beta_2 W y + \varepsilon$$

where the dependent variable y_c is the relevant contract item (such as National Board certification bonus or number of salary columns), X_i' is the vector of explanatory variables, $\beta_2 W y$ is amount that outcomes are affected by nearby districts, and ε is the error term. The naïve estimates do not include the $\beta_2 W y$ term, and I use probit regression where appropriate (for binary rather than continuous indicators).

I pause here to explain why I believe that, in addition to significant Moran test results for spatial dependence, the incorporation of spatial weights is crucial even if they do not significantly improve the fit of the overall models. Though individual teachers or new members to the school board may not know the ins and outs of their union contract, at least a few key individuals must. At a minimum, the district's legal counsel must be cognizant of neighboring districts – in particular, the rights and the benefits offered to both new and veteran hires. Considering the sameness of many contracts, no doubt there is sharing among districts in the same region or among unions in the same intermediary organization. The static nature of contracts also speaks to institutional knowledge, where contracts are allowed to remain static because both parties assume they are already receiving the best possible compromise. I am using all neighboring districts for

comparison as a reasonable distance that potential employees might travel for higher pay, which is a well-documented phenomenon (e.g., Associated Press, 2019; Murnane, Singer, Kemple, & Olson, 2009).

For research questions 3 and 4, I look at contracts as a whole, consistent with the existing literature (Strunk et al., 2018). In order to use contract strength as a "treatment" (and use less strong contracts as a "control"), I create a dichotomous variable using a cut-off point based on the distribution of contract strength within the state. I also test a categorical variable with either moderately-strong or least-strong contracts as the control group, in order to examine the sensitivity of the cut-off point. These cut-offs points occurred at 0.25 and 0.50 after examining the distribution of the data. I discuss the decision-making behind these cut-offs in more depth in Chapter 5.

I also treat each high-profile provision (such as National Board Certification and guaranteed prep time) as binary treatment variables. With a strength measure created, I can then leverage a propensity score weighting (PSM) approach (Rosenbaum & Rubin, 1983). PSW allows me to compare districts that are most similar to each other across a range of characteristics, with the exception of the strength of the contract (or, the specific high-profile provision). Despite its many limitations, PSW is the most appropriate technique for this study considering that n=499, and I am working with one year of data. The equation for obtaining the propensity score is as follows:

$$D(X_i) = \Pr(RES_i = 1 \mid X_i)$$

where $D(X_i)$ is the propensity score given a set of predictors X_i. Among these predictors, I include independent variables for enrollment well as per pupil expenditures, enrollment

for several student subgroups (low-income, students on IEPs, and ELs), and partisanship (voters and school boards). Because of the Pennsylvania School Data Project (RFA, 2018), I am also able to include additional independent variables such as the average years of teacher experience in each school district.

To obtain school board partisanship, I use the MatchIt command in Stata to match school board members to publicly available voter registration files with 70-80% success depending upon the matching parameters (Meredith & Grissom (2010) in a similar study were only able to obtain 80% accuracy). There is no centralized data collection of the nearly 5,000 school directors in Pennsylvania, and directors make use of nicknames or exclude middle names, which hampers the accuracy of the MatchIt command. Furthermore, the overwhelming majority of directors cross-file under both parties, except for the least popular candidate in a handful of districts. Throughout Chapters 4 and 5, I compare and contrast both voter registration and school board composition and find few differences, though the school board estimates are noisier and less precise than voter registration. Because of these limitations, I consider the voter registration of a school district to be the more accurate measure, as directors may be more responsive to constituent demands during contract negotiations, irrespective of party.

Using $D(X_i)$, I can then obtain a set of weights to estimate the average treatment effect on the treated as follows: $w(x) = \frac{D(X_i)}{1-D(X_i)}$. These weights render the following property, $f(x|PPL = 1) = f(x|PPL = 0, w(x))$, which means that the distribution of the characteristics of districts with the strongest contracts will be statistically the same as the

distribution of the characteristics of districts with the weakest contracts. I rely on weighted linear probability models using the $w(x)$ described above in the form:

$$y_i = a + \beta_1 + \beta_i X_i' + \varepsilon | w(x)$$

where y_i is student achievement (scores on the relevant state standardized test), X_i' is the vector of explanatory variables after weighting to conduct doubly robust estimations, β_1 is the coefficient of interest associated with the treatment effect of strong contracts, and ε is the error term.

I obtain the propensity scores ($D(X_i)$) through the use of machine-learning boosted models to reduce overfitting, known as a generalized boosted models (GBM) (González Canché, 2018; Ridgeway, McCaffrey, Morral, Burgette, & Griffin, 2014). Essentially, this procedure runs thousands of iterations in order to achieve more stable and less biased results. It selects results with optimal balance and is more precise than traditional PSW procedures, though with a single year of data it would still be inappropriate to make causal claims.

Finally, I conduct subgroup analysis, as there may be different relationships between strength and achievement among student subgroups in districts. Throughout Chapters 4 and 5, I present the results for all students and all other groups when available. I hypothesize, in keeping with Lovenheim and Willén (2019), that stronger CBAs may negatively affect historically disadvantaged students. Pennsylvania has avoided reporting data on racial subgroups and instead uses a historically disadvantaged super subgroup and, only in recent years, an economically disadvantaged group (see Wrabel, Saultz, Polikoff, McEachin, & Duque, 2018, for the political rationale behind this method of data

reporting). I aim to provide a nuanced and comprehensive analysis by offering these effect heterogeneity tests, though I am only able to do so for the historically disadvantaged group because of the limitations of the data.

Before presenting my findings in Chapter 4, I summarize how my theoretical framework drives my methodological choices. The framework is unidirectional, leading from characteristics to items to student outcomes (either achievement or graduation rates). It accounts for structural differences both in terms of the policy context and in terms of nearest neighbors, which necessitates the inclusion of spatial matrices and weights. The predictive research questions – determining whether partisan factors predict contracts and contract items – demand the inclusion of neighboring characteristics as well as each district's own demographics. Finally, propensity score weighting allows me to compare the spatial autoregression approach (GS2SLS) with the GBM method, both for fit and for patterns of significance. I maintain that results that hold across both methods are most persuasive. As we will see, only one significant finding holds across all of these different methods, and contract items may be best predicted by whether they exist in neighboring districts.

CHAPTER 4: VARIATION AND RELATIONSHIPS IN COLLECTIVE BARGAINING AGREEMENTS

This chapter refers back to Table 1 and illustrates, using QGIS mapping software, the distribution of CBA items across Pennsylvania. At the end of each Panel, I present the results of spatial regression models to describe the likelihood that an item will appear in a district. I consider the following variables in every model: total per pupil expenditures, total enrollment, percentage of low-income students, percentage of students on individualized education plans (IEPs), percentage of limited English proficient students (LEP), student-to-teacher ratio, percentage of dollars obtained from local revenues (as opposed to state or federal), and percentage of registered Republicans. Table 2 describes the mean, standard deviation, minimum, and maximum values for the independent variables in my models.

Table 2. Characteristics of Pennsylvania School Districts

	PPE	Enroll	% LI	% IEP	% ELL	S-T Ratio	% Local	% Rep
Mean	$17,398	3,163	43	18	1	14	52	46
SD	($3,354)	(6,468)	(17)	(4)	(3)	(2)	(19)	(13)
Minimum	$12,020	199	4.5	8	0	8	10	7
Maximum	$37,825	134,129	100	36	24	20	86	73

"PPE" refers to per-pupil expenditures, "% LI"refers to percentage of low-income, "% IEP" refers to percentage of students on individualized education plans, "% ELL" refers to percentage of English learners, "S-T" is the student-teacher ratio, "% Local" is the percentage of expenditures that come from local property taxes, and "% Rep" is the percentage of Republicans in a district.

Because board composition is strongly and positively correlated with voter registration, I opt to use the percentage of registered Republicans in the analysis below; it would not be appropriate to include both highly correlated (Pearson's r=0.70) variables

in the models. Voter registration is a more precise and nuanced measure, and it seems unlikely that board members would differ sharply from that of their community, particularly in small rural districts. This approach is also a better fit for including Philadelphia in the data, which until recently operated under a five-person School Reform Commission (SRC). The governor appointed SRC members, reflecting the governor's political party rather than that of Philadelphia, which is heavily Democratic (and only 11.24% Republican). The SRC disbanded itself in 2017 (Graham, 2017). Party registration in the aggregate may be a good indicator of party preferences, but board members' individual registration may not adequately reflect their stance on local issues of school governance (Meredith & Grissom, 2010).

Salary and Salary Structures

First, I consider the primary salary structures in Pennsylvania, including the number of columns, the number of rows, and the minimum and maximum salary. Pennsylvania during the 2015-2016 school year had an average teacher salary of $64,447 and ranked tenth highest in the nation for teacher pay (NEA, 2016). Higher teacher salaries are concentrated in the suburbs as shown in Figure 4. Figure 5 adjusts salaries using the comparable wage price index, which is a measure of wages for similar college graduates in the district (Taylor, 2006).

Figure 4. Average Salary Quintiles

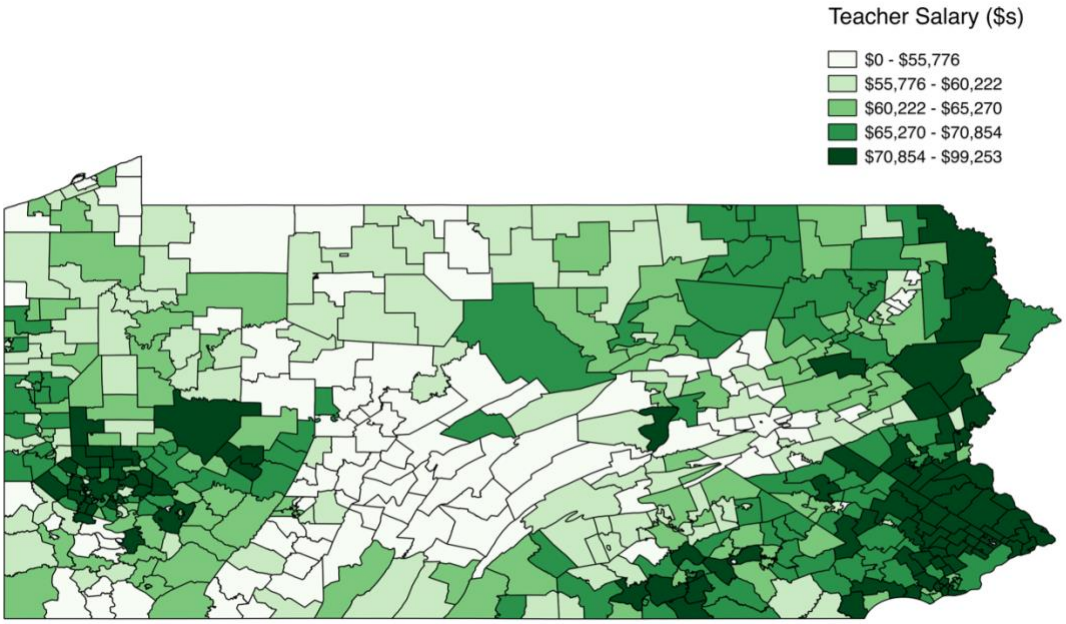
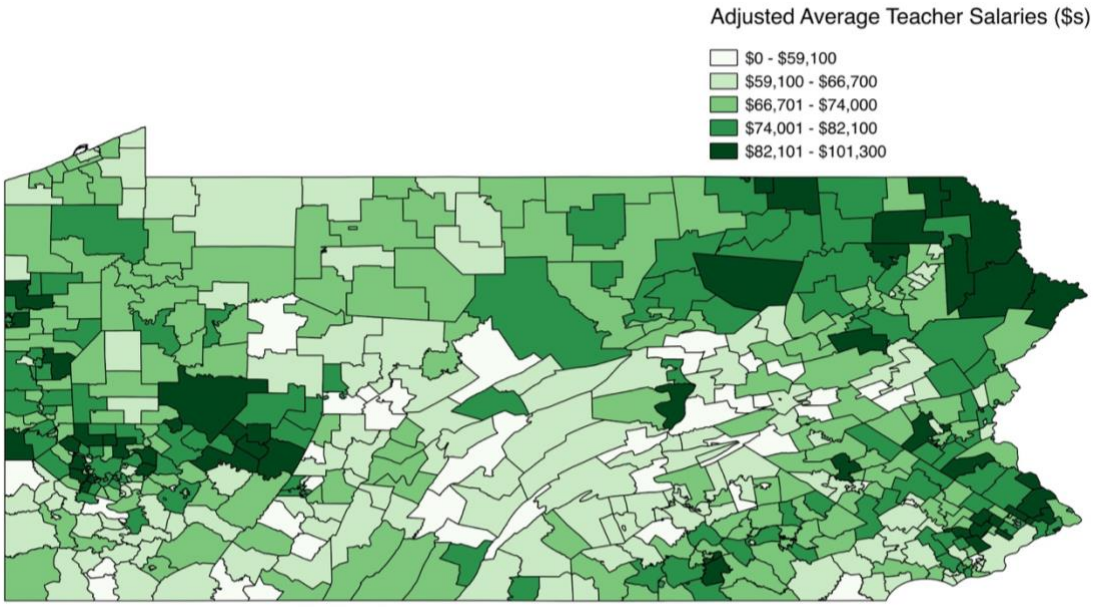


Figure 5. Average Adjusted Salary



Teacher salaries are perhaps the most obvious outcome of a collective bargaining agreement. Thus, it is important to understand the factors that predict them, including

geospatial factors, in order to test other models for more specific contract items. Significant spatial dependency would suggest that districts are aware of competing salaries in the labor market.

After considering several models, I included interaction variables with urban districts, which improved the fit of the model. After the first OLS regression shown in the 1st column of Table 3, I performed the Moran test for spatial correlation among the residuals, which was significant ($\chi^2 = 104.26$, p > 0.00). This finding suggests awareness of neighboring salaries. In columns 2 and 3, I show the results of both the maximum likelihood model with robust clustered standard errors and the generalized spatial two-stage least squares (GS2SLS). In columns 4-6, I also perform the same tests for the adjusted teacher wages (see: Taylor (2006) for this procedure, which relies on regional wage adjustments using the American Community Survey (ACS) index).

Table 3. Regression Results for Average and Adjusted Teacher Salaries by District, 2016-2017 School Year

	1	2	3	4	5	6
	Unadjusted $, OLS	Unadjusted $, Spatial ML	Unadjusted $, Spatial GS2SLS	Adjusted $, OLS	Adjusted $, Spatial ML	Adjusted $, Spatial GS2SLS
PPE ($s)	1.170**	0.860**	0.898**	0.987**	0.783**	0.790**
	(0.108)	(0.096)	(0.098)	(0.128)	(0.109)	(0.110)
Enrollment	0.872**	0.530**	0.552**	0.324	0.436*	0.378*
	(0.169)	(0.164)	(0.166)	(0.199)	(0.185)	(0.186)
% Low-Inc.	-82.589**	-88.730**	-87.703**	-21.913	-62.757*	-57.680
	(29.370)	(26.972)	(27.300)	(34.606)	(30.546)	(30.803)
% IEP	-10.706	-16.210	-14.062	-283.812*	-161.942	-177.557
	(97.274)	(88.729)	(89.965)	(114.617)	(100.465)	(101.401)
% ELL	74.950	184.801	193.207	-552.011**	-144.715	-159.849
	(167.578)	(151.109)	(153.523)	(197.455)	(171.017)	(173.172)
S-T Ratio	1,391.283**	1,106.492**	1,139.490**	1,314.792**	1,080.357**	1,093.775**
	(226.617)	(197.772)	(201.800)	(267.020)	(223.742)	(226.927)
% Local Tax	163.613**	127.779**	137.289**	63.227*	69.447*	67.668*
	(25.631)	(25.093)	(25.035)	(30.200)	(28.491)	(28.467)
% Republican	-98.277**	-64.418*	-77.152**	-33.906	2.845	-4.639
	(25.277)	(30.503)	(29.451)	(29.784)	(34.879)	(34.211)
Urban	-5,329.591	1,185.510	121.326	-6,593.959	1,369.494	809.731
	(5,769.562)	(4,899.400)	(5,015.085)	(6,798.213)	(5,539.685)	(5,629.520)

Urban x Enroll	-0.938**	-0.726**	-0.740**	-0.457*	-0.578**	-0.533**
	(0.175)	(0.158)	(0.161)	(0.206)	(0.179)	(0.180)
Urban x Low-I	122.044	31.390	43.519	179.862	44.700	49.960
	(90.401)	(76.916)	(78.759)	(106.519)	(86.945)	(88.368)
constant	21,390.805**	30,744.932**	29,217.982**	38,861.073**	40,751.353**	40,660.559**
	(5,386.304)	(4,952.980)	(5,009.846)	(6,346.624)	(5,611.098)	(5,658.500)
N	499	499	499	499	499	499
R^2 / *Pseudo-* R^2	0.589	0.571	0.574	0.254	0.234	0.238

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Scanning across Table 3 highlights the importance of testing for spatial factors and regional wage differences considering the changes in the magnitude of the coefficients of interest. In the unadjusted models (columns 1-3), the percentage of registered Republican voters in a district significantly and negatively predicts salary – each percentage point of registered Republicans in the district translates into a $64-77 decrease in teacher pay. But this relationship becomes nonsignificant in the adjusted salary models. Controlling for spatial differences also changes the magnitude of the relationship between per-pupil expenditures (PPE) and average teacher salary – every additional dollar predicts a minimum of a 0.78 cent increase (column 5). This relationship is larger in the unadjusted columns (2 and 3) – 0.86 or 0.90 cent increase in the average salary.

Enrollment does increase teacher salaries, but these gains are offset in urban areas in the models without spatial controls (columns 1 and 5). This finding confirms a long history of inequitable funding in the state (Steinberg & Quinn, 2015). Additionally, districts with higher proportions of low-income students pay teachers less in the unadjusted models, though this negative relationship shrinks in the adjusted models and becomes nonsignificant, from an $83 decrease in column 1 to a $58 decrease in column 6.

Across every model, teachers with a higher student-teacher ratio are earning between $1,080-$1,391 more for every increase in the ratio. This finding matches a small body of literature highlighting that class size reduction measures can lead to teacher salary decreases in order to budget for additional hires (e.g., Hanushek & Luque, 2000;

Hoxby, 2000b). This relationship shrinks when adding spatial controls. Predictably, a higher percentage of local property taxes funding a district's schools also translates into higher teacher pay, though this too shrinks too from a $164 unadjusted average salary increase to a $68 adjusted average salary increase in column 6. Districts have a variety of expenditures (capital, debt, instructional materials, administrative salaries and more) in addition to teacher salaries.

In light of these procedures, I consider column 6 to offer the most robust estimates. Looking only at column 6, the significant positive predictors for adjusted average teacher pay in a district are PPE, enrollment, student-teacher ratios, and the percentage of a district's budget that comes from local property taxes. An increase in any of these measures predicts an increase in salary. The only significant negative predictor is the interaction between urban districts and enrollment. Without controlling for spatial differences, the partisan finding would be more meaningful – but I show that salaries are better predicted by proximity than partisanship. Keeping salaries competitive with neighboring districts is not a partisan issue.

I turn next to the structure of the salary schedule, a CBA feature used by 99% of districts. Where districts did not use salary schedules (i.e., they awarded no set bonuses for additional degrees or graduate credits), I inputted a "1" for number of rows (representing only 7 small rural districts). I verified those 7 districts that did not have a salary scale attached to their collective bargaining agreement by making phone calls and sending emails to confirm that an appendix had not been excluded from the Right-to-

Know (RTK) requests. Figure 6 demonstrates the extent to which districts rewards teachers for years of experience, with each row representing a year of experience.

Rural districts, based on Figure 6, seem to value experience more heavily than urban counterparts, which may have as few as 10 salary rows. However, the point-biserial correlation coefficient between rural and rows is not significant (r = -0.08).

Figure 6. Number of Rows in the Salary Schedule

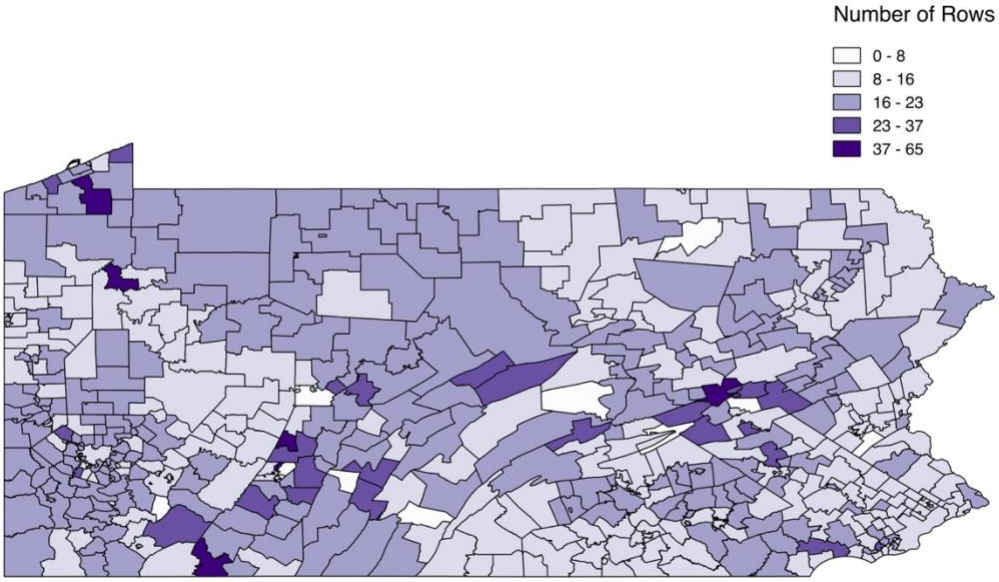


Figure 7 reveals that the use of multiple credential-based columns is more prevalent in the eastern part of the state, as well in some of the suburbs near Pittsburgh in the western half. The point-biserial correlation coefficient is significant here between rural districts and columns (r = -0.10) and between suburban districts and columns (r = 0.15). Districts with more columns are incentivizing the earning of additional graduate degrees and credentials (though they may or may not provide tuition remission to do so). Philadelphia and Pittsburgh are more similar to some of the rural districts because of their unique merit pay structures that followed Race to the Top-era incentives (Bleiberg &

Harbatkin, 2018). Erie City and Troy Area school districts have the greatest number of columns at 18; they award salary increases for every 6 credits earned (starting from column 1, Bachelor's, through column 18, Doctorate.)

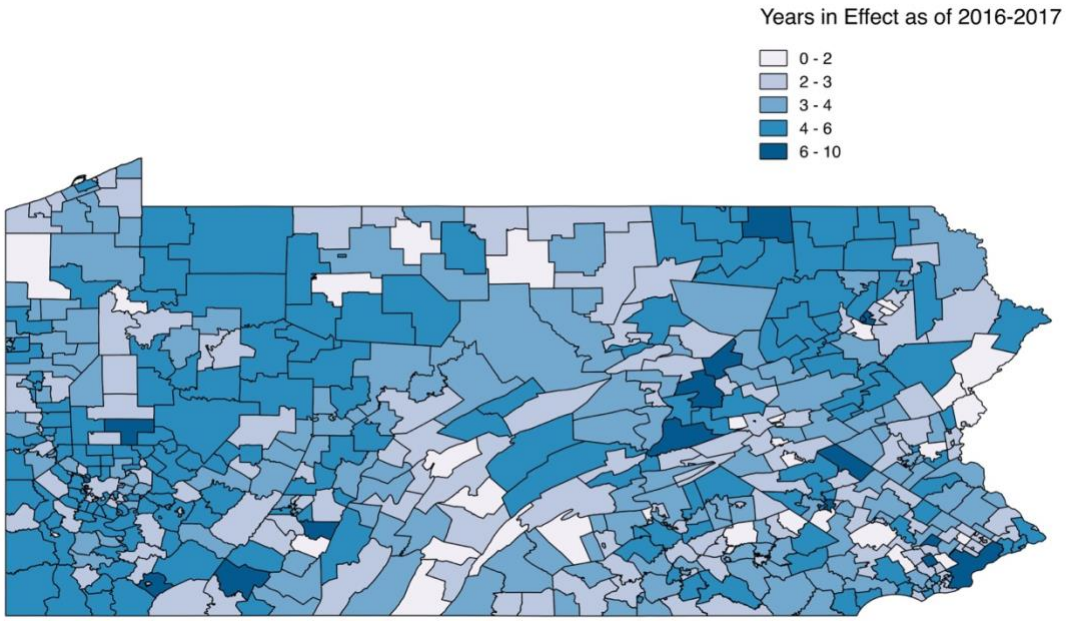
Figure 7. Number of Columns in the Salary Schedule

Figure 8 demonstrates that long contracts are concentrated in urban and suburban districts (r = 0.19), whereas rural districts rely on relatively short contracts (r = 0.20). There is a strong leftward skew to the data with an average length of 48.41 pages (*SD* 19.84 pages). As for the duration of the contracts, there is a leftward skew with an average duration of 4 years (*SD* 1 year).

Figure 9 demonstrates that contracts in effect for longer periods of time are distributed throughout the state, and there are no significant point-biserial correlations among the three district types.

Figure 8. CBA Length

Figure 9. CBA Duration



Looking at all of these continuous variables, I ran Moran tests to examine spatial dependence. The Moran test was significant for columns ($\chi^2 = 77.65$, $p < 0.00$) and

length ($\chi^2 = 82.26$, p < 0.00). It was not significant for rows ($\chi^2 = 2.38, \mathrm{p} < 0.05$) and years in duration ($\chi^2 = 3.75$, p < 0.01). I present the results of these regressions in Table 4, using either OLS regression or spatial autoregression with GS2SLS, dependent upon whether the χ^2 statistic was significant. It is noteworthy that districts are significantly more likely to include more graduate credit bonuses (columns) if there neighboring district does so. Considering the resistance to adopting Race to the Top reforms, adjacent district pressures may be stronger than isomorphic pressures or diffusion from urban areas.

Table 4. OLS and Spatial Autoregression Results (GS2SLS) for Contract Features

	Rows (OLS)	Columns (GS2SLS)	Length (GS2SLS)	Years (OLS)
PPE ($100ss)	-0.170	0.015	0.111	0.023
	(0.088)	(0.041)	(0.277)	(0.021)
Enroll (1000s)	-0.106	0.099	1.233**	-0.035
	(0.138)	(0.068)	(0.455)	(0.033)
% Low-Inc.	-0.023	0.000	0.026	-0.003
	(0.024)	(0.011)	(0.076)	(0.006)
% IEP	-0.024	0.071	0.364	-0.005
	(0.080)	(0.038)	(0.253)	(0.019)
% ELL	-0.002	0.056	-0.437	0.007
	(0.138)	(0.064)	(0.435)	(0.033)
S-T Ratio	-0.465*	-0.025	0.014	0.075
	(0.185)	(0.084)	(0.574)	(0.044)
% Local Tax	-0.031	0.036**	0.109	-0.005
	(0.021)	(0.010)	(0.068)	(0.005)
% Republican	-0.332	-0.014	-2.024	-0.258
	(0.675)	(0.326)	(2.168)	(0.163)
Urban	1.168	-2.869	20.642	-0.326
	(4.749)	(2.115)	(14.502)	(1.143)
Urban x Enroll	0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Urban x Low-I	-0.031	0.055	-0.107	-0.003
	(0.074)	(0.033)	(0.228)	(0.018)
_cons	30.377**	2.556	21.412	3.342**
	(4.048)	(1.888)	(12.769)	(0.974)

N	499	499	499	499
R^2 / *Pseudo-*R^2	0.045	0.138	0.194	0.027

There are a few significant predictors for the continuous dependent variables in Table 4. Higher PPE does not significantly predict any of these four contract features. Increasing the student-teacher ratio does significantly predict fewer rows (rewards for experience), with an increase of 2 students (from 18:1 to 20:1, for example) predicting a decrease of one row. The low r^2 value, however, should caution against overinterpretation, and later chapters will use doubly robust PSW techniques.

As for columns, the percentage of local property tax revenue paying for schools is a significant positive predictor; wealthier districts may place a higher value on teacher credentials. There is one other significant predictor, for length – enrollment, which is not surprising, considering that the high enrollment districts of Philadelphia and Pittsburgh have the longest contracts. The most compelling finding from this correlational analysis is the relationship between tax revenue and columns, as districts more reliant on federal and state funding may be less willing to pay for teachers to earn additional credits and degrees – and more willing if neighboring districts also provide these incentives.

Bonuses and Class Size

Only six school districts offer hiring bonuses for hard-to-recruit teachers, and these bonuses were all for special educators. This bonus ranged from $500 to $2,000. These districts are Northeast Bradford, Northern Lebanon, Berwick Area, Wyalusing Area, Connellsville Area, and Wyoming Valley West, as shown in Figure 10.

Figure 10. Districts with a Special Educator Bonus

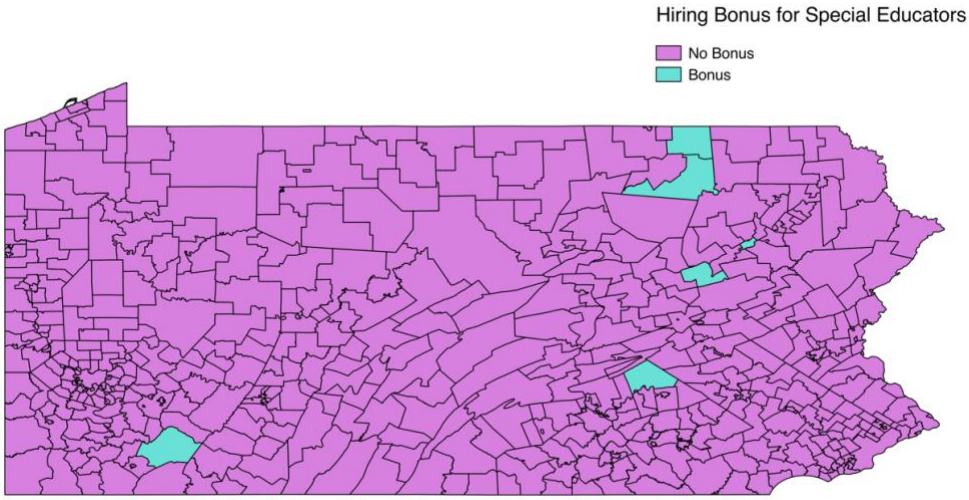


Figure 11 demonstrates the prevalence of longevity bonuses, and Figure 12 demonstrates the prevalence of NBC bonuses. Figures 13-15 show the relative rarity of class size restrictions in the state, which appear in urban (r = 0.09) and suburban districts (r = 0.13). The point-biserial correlation is negative and significant for rural districts (r = -0.12)

Figure 11. Longevity Bonuses

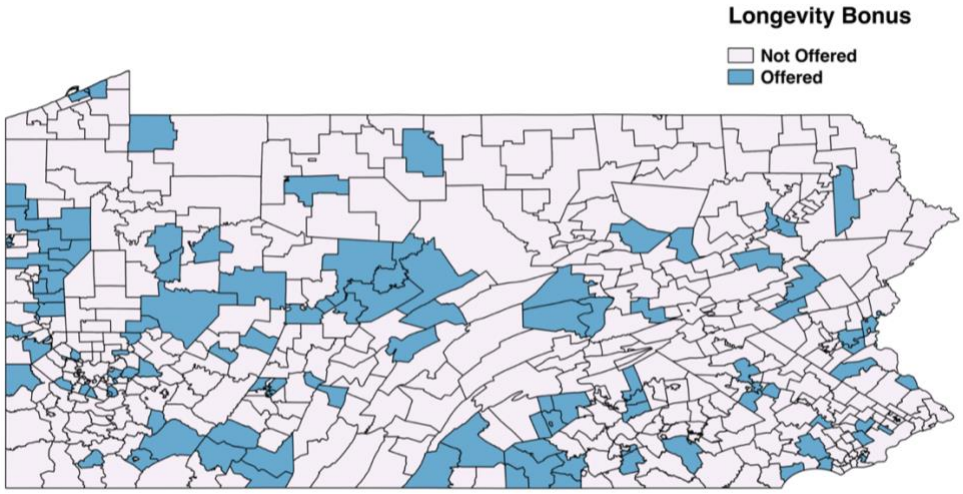


Figure 12. National Board Certification Bonuses

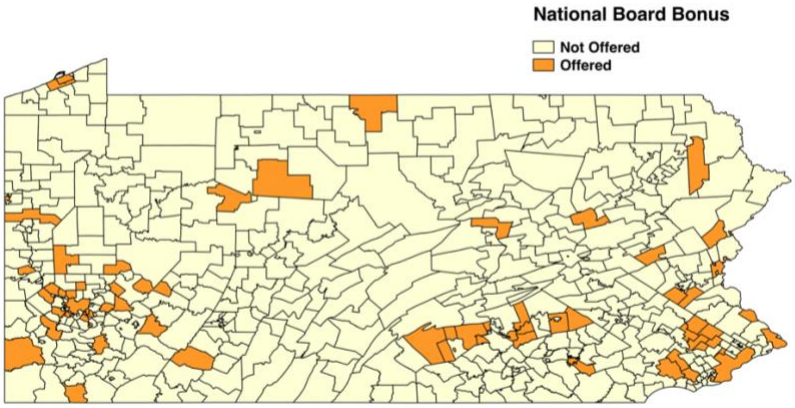


Figure 13. Fourth Grade Maximum Class Sizes

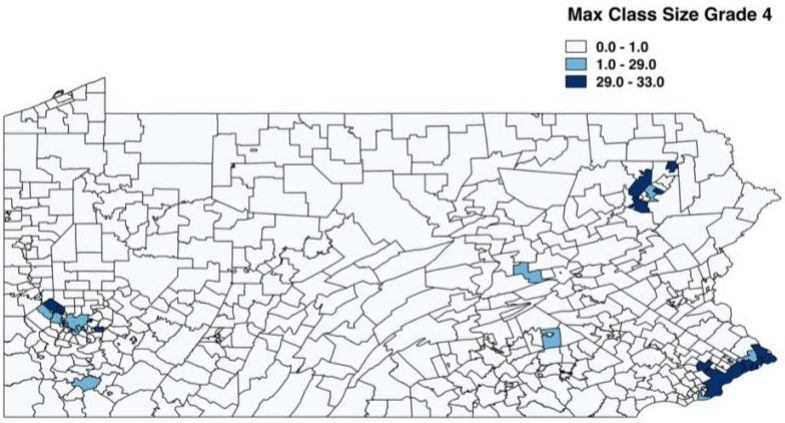


Figure 14. Eight Grade Maximum Class Sizes

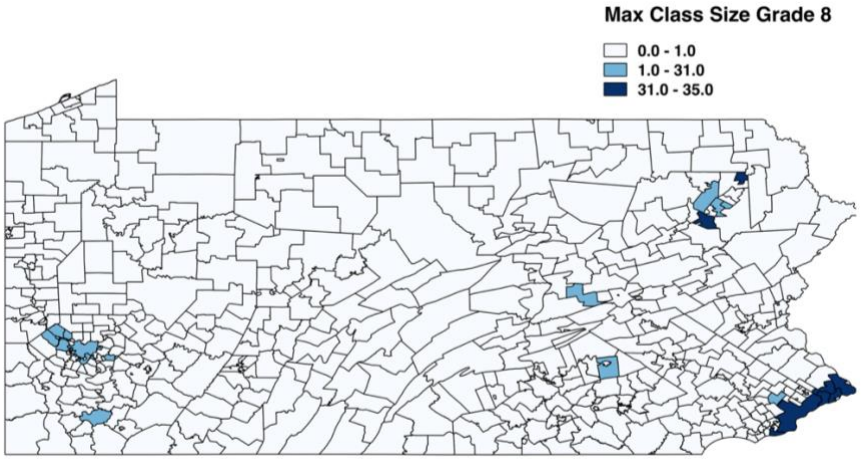
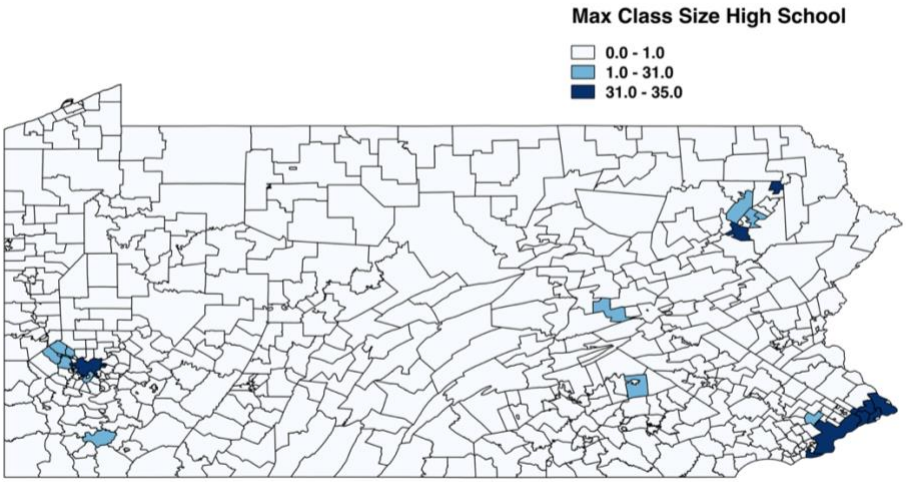


Figure 15. High School Maximum Class Sizes



In Table 5, I use probit models, and I examine whether a district has a maximum class size cap instead of the continuous variable for class size restrictions. I report the marginal effects. I omitted the town variable because there were 0 town districts with some of the relevant contract items. I compared probit to logit models, but the logit models provided much noisier estimates. I elected to use Republican school boards for ease of interpretation (instead of registered Republicans as a continuous variable), though the relationships are similar in magnitude and in the same direction. I use whether a district has a majority Republican school board in order to predict the presence of these binary outcomes. Table 5 illustrates that Republican school boards are significantly less likely to restrict class sizes, and that each percentage increase of local property tax percentage increases the likelihood that a district will enforce class size maximums.

Table 5. Probit Models for Bonuses and Maximum Class Sizes

	Bonus for SPED	Longevity Bonus	NBC	Max Class 4th	Max Class 8th	Max Class High
PPE ($100ss)	-0.001	-0.002	-0.000	-0.002	-0.002	-0.002
	(0.001)	(0.007)	(0.005)	(0.002)	(0.002)	(0.002)
Enroll (1000s)	0.000	-0.001	0.014*	0.003	0.003	0.003
	(0.000)	(0.004)	(0.006)	(0.002)	(0.002)	(0.002)
% Low-Inc.	0.000	0.002	-0.000	0.001	0.001	0.001
	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
% IEP	0.001	-0.013*	-0.007	0.002	0.002	0.002
	(0.001)	(0.006)	(0.005)	(0.002)	(0.002)	(0.002)
% ELL	-0.001	-0.000	0.001	-0.002	-0.002	-0.002
	(0.001)	(0.009)	(0.007)	(0.003)	(0.003)	(0.003)
S-T Ratio	0.002	-0.020	-0.023*	-0.002	-0.002	-0.002
	(0.002)	(0.015)	(0.012)	(0.005)	(0.005)	(0.005)
% Local Tax	0.000	-0.000	0.002	0.001*	0.001*	0.001*
	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Rep. Board	-0.004	-0.096	-0.044	-0.057**	-0.057**	-0.055**
	(0.005)	(0.053)	(0.040)	(0.018)	(0.018)	(0.018)
Rural	0.008	-0.016	-0.032	-0.000	-0.000	0.001
	(0.010)	(0.054)	(0.045)	(0.024)	(0.024)	(0.024)
Suburban	-0.004	-0.029	0.045	0.017	0.017	0.015

	(0.006)	(0.056)	(0.046)	(0.022)	(0.022)	(0.022)
_cons	-0.001	-0.002	-0.000	-0.002	-0.002	-0.002
	(0.001)	(0.007)	(0.005)	(0.002)	(0.002)	(0.002)
N	499	499	499	499	499	499
pseudo R^2	0.197	0.019	0.123	0.225	0.225	0.222

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Teacher Evaluation, Layoffs, and Rehiring

Referring back to Table 1, I depict the items that are variable in Figures 16-20. Remember that many of the provisions in the existing CBA literature (see: Strunk et al., 2018) are not relevant to Pennsylvania because of state mandates. The presence of these items in the contract also does not necessarily mean that these provisions are not in effect because the Pennsylvania School Code allows for seniority-based hiring, firing, and re-hiring without additional contract negotiation needed, though this changed in 2017 (PDE, 2018) after my data collection. This policy change is an important area for future study. Recent changes have superseded many of these provisions by allowing for teachers to be laid off in terms of effectiveness, not seniority.

In Figure 16, the most common consequence for negative evaluation scores was not advancing on the salary scale, or termination after two semesters without improvement. As a reminder, the overwhelming majority of teachers in Pennsylvania receive satisfactory ratings (Steinberg & Kraft, 2017). Figure 17 depicts those districts that specifically allow for an evaluation rebuttal process separate from the standard grievance and complaint procedures. Figure 18 shows those districts that require a written growth plan for teachers who receive negative evaluations, while Figure 19 shows those districts that explicitly specify seniority-based layoffs. This specificity is irrespective of overlap with existing provisions in the Pennsylvania School Code, which provided for seniority-based layoffs and recalls in Act 97 (passed in 1979) and was only recently changed in 2018 to use teacher evaluation ratings in these decisions (PDE, 2018). Last,

Figure 20 demonstrates which districts require that the most senior employees must be re-hired first after layoffs that result from economic hardship.

Figure 16. Districts with Defined Consequences for Negative Evaluations

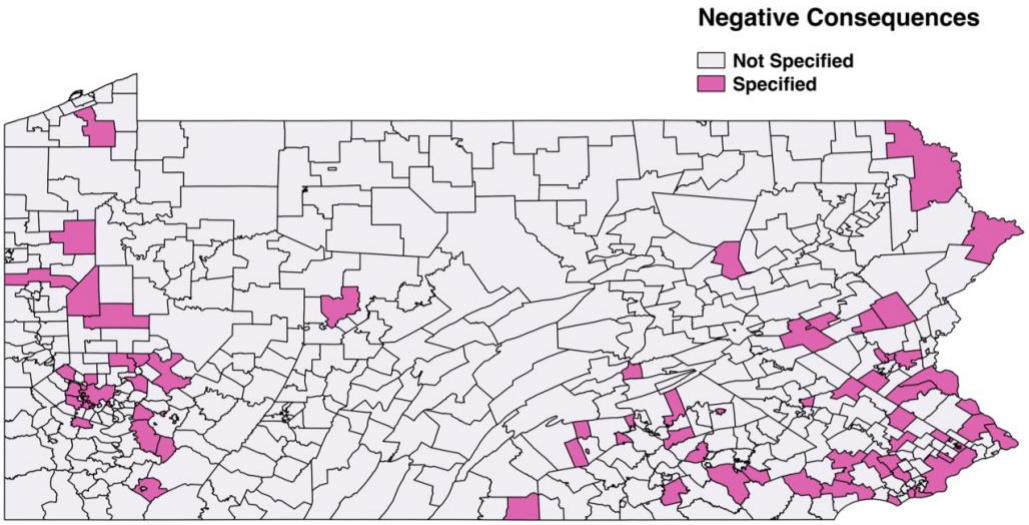


Figure 17. Districts with Negative Evaluation Rebuttal Provisions

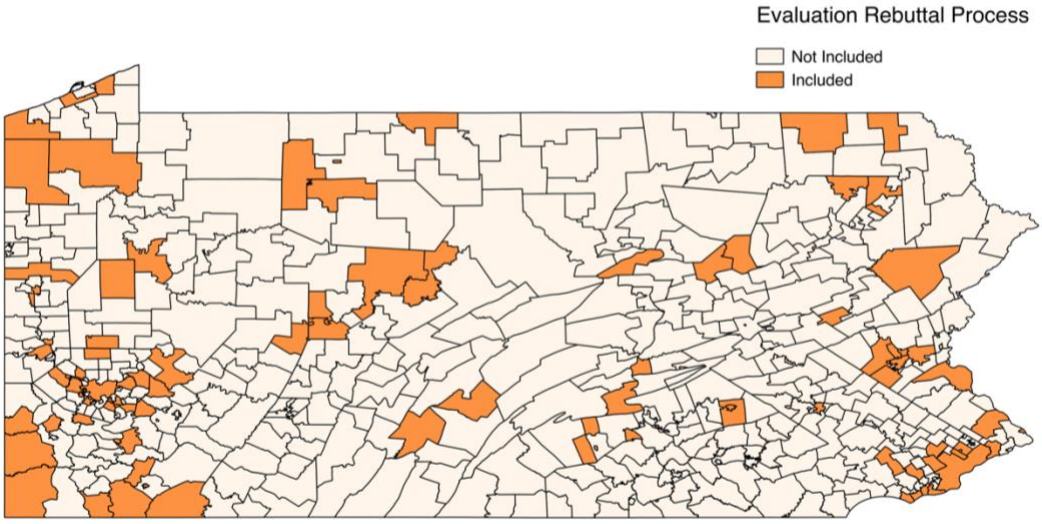


Figure 18. Districts with Mandated Professional Growth Plans

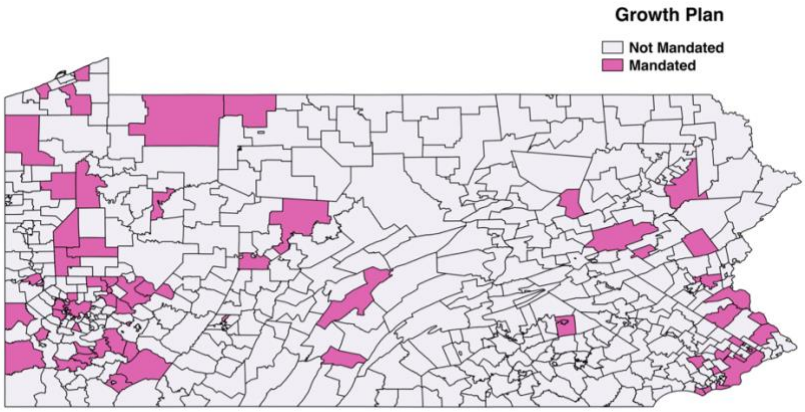


Figure 19. Districts with Seniority-Based Layoffs

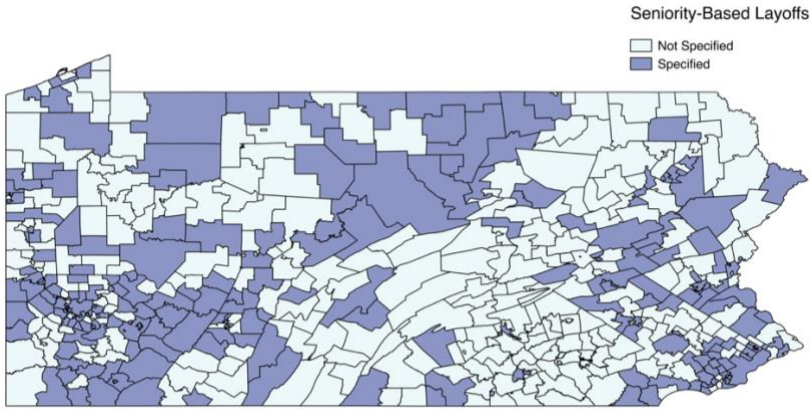


Figure 20. Districts with Seniority-Based Recall

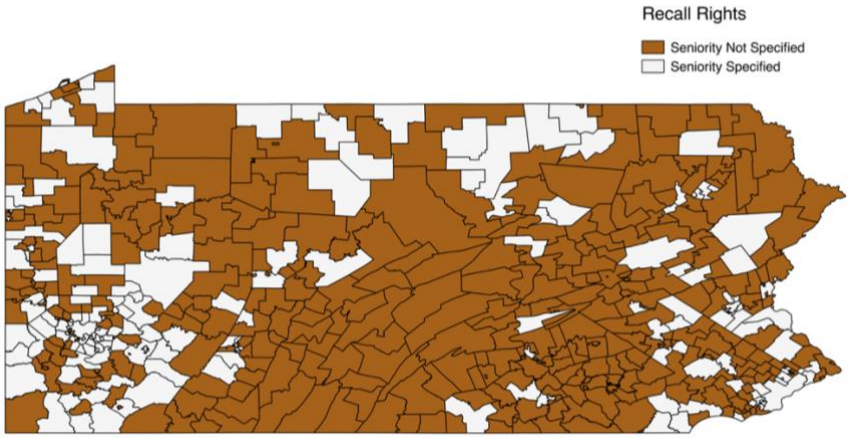


Table 6 uses probit models to predict the presence of these items. As the suburban variable was significant in these models, I interacted it with items that I theorized might influence decision-making differently in wealthier suburban districts, and which were significant prior to testing for interactions. The inclusion of these interaction variables significantly improved the fit of the models. I also performed the Moran test for each of the items: ($\chi^2 = 0.81$ for consequences, 1.00 for rebuttal, and 2.59 for growth plans, 9.88 for seniority layoffs, and 11.98 for recall rights). The final two items are spatially dependent – districts that had strong seniority protections were more likely to have neighboring districts with these same protections. Here again is more evidence that districts are aware of their neighbors and respond accordingly.

Despite the inclusion of interaction terms, looking at the pseudo R^2 statistic, none of these models are a particularly good fit for the data. There are no significant predictors for negative consequences. Urban districts with higher per pupil expenditure are significantly more likely to provide a distinct evaluation rebuttal process, while suburban and town districts are significantly less likely. Districts with Republican school boards are significantly less likely to mandate a professional growth plan. Districts with higher per pupil spending are significantly more likely to specify seniority-based layoffs and recall rights, and these recall rights are significantly less likely to appear in districts governed by Republican school boards. These findings are consistent with the notion that Republicans prefer fewer seniority rights and fewer process-based alternatives (observations) compared to stringent, test-based evaluation systems (Tetlock et al., 2013).

Table 6. Probit Models for Teacher Evaluation Consequences, Layoffs, and Rehiring

	Negative Conseq.	Rebuttal	Growth Plan	Seniority Layoff	Seniority Recall
PPE ($100ss)	0.003	0.015*	0.005	0.017*	0.021*
	(0.005)	(0.006)	(0.005)	(0.009)	(0.008)
Enroll (1000s)	0.008	0.014	0.008	0.012	0.019
	(0.006)	(0.008)	(0.007)	(0.011)	(0.011)
% Low-Inc.	-0.003	0.000	-0.001	-0.008*	-0.006
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
% IEP	0.009	0.005	0.008	0.009	0.010
	(0.005)	(0.006)	(0.005)	(0.008)	(0.007)
% ELL	-0.003	-0.010	0.001	0.023	0.008
	(0.008)	(0.009)	(0.008)	(0.013)	(0.011)
S-T Ratio	0.010	-0.011	0.002	-0.042	-0.013
	(0.015)	(0.016)	(0.015)	(0.023)	(0.022)
% Local Tax	0.002	-0.005*	0.001	-0.009**	-0.007*
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Rep. Board	-0.001	-0.059	-0.089*	-0.114	-0.149*
	(0.042)	(0.047)	(0.043)	(0.067)	(0.063)
Rural	-0.129	-0.118	0.070	-0.106	-0.006
	(0.076)	(0.096)	(0.135)	(0.170)	(0.173)
Suburban	-0.019	-0.826**	0.395	-0.336	-0.281
	(0.404)	(0.245)	(0.490)	(0.607)	(0.574)
Town	-0.074	-0.160**	-0.009	-0.028	0.032
	(0.065)	(0.049)	(0.117)	(0.171)	(0.177)
Sub x Local	0.000	0.006*	-0.004	0.005	0.000
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Sub x Low	0.002	0.001	-0.002	0.006	0.004

	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Sub x Ratio	-0.012	0.038	0.003	-0.003	0.021
	(0.019)	(0.022)	(0.020)	(0.032)	(0.029)
_cons	0.003	0.015*	0.005	0.017*	0.021*
	(0.005)	(0.006)	(0.005)	(0.009)	(0.008)
N	499	499	499	499	499
pseudo R^2	0.116	0.093	0.056	0.070	0.078

School Day

Teacher prep time remains an issue of contention between unions and administrators, as high-need schools may be much more likely to infringe on teacher prep time to assign additional duties such as lunchroom supervision, last-minute parent conferences, and special assemblies. Contractually negotiated prep time can also be used for professional learning communities and other activities, which may help better align instruction and improve student achievement. Missed opportunities to prepare creative, innovative lessons may "accumulate over time, undoubtedly affecting student learning" (Merritt, 2017, p. 31). As Figures 21-23 show, however, some teachers in Pennsylvania have no guaranteed planning time at all, while others have far more than the average. There seems to be a "hollowing out" of prep time in central Pennsylvania.

Figure 21. Fourth Grade Guaranteed Preparation Time

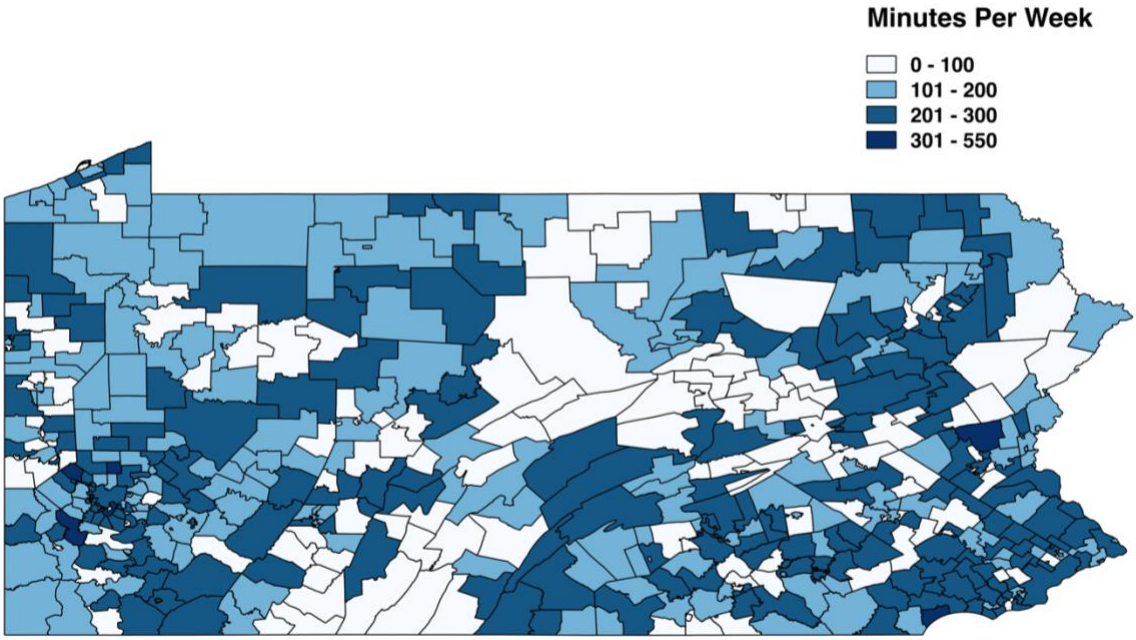


Figure 22. Eighth Grade Guaranteed Preparation Time

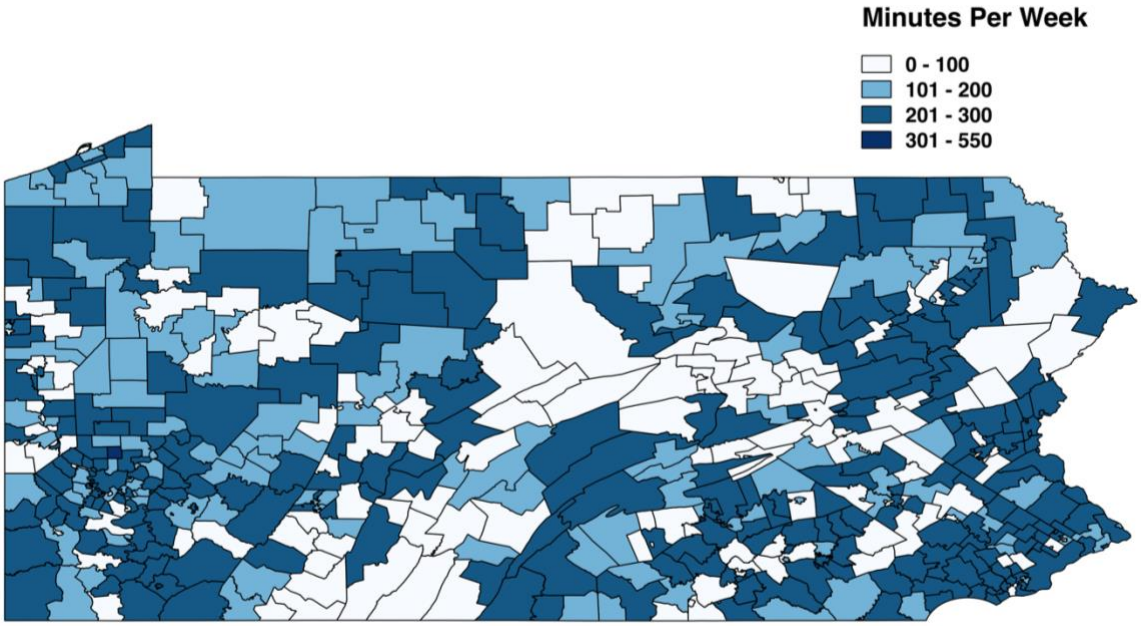


Figure 23. High School Guaranteed Preparation Time

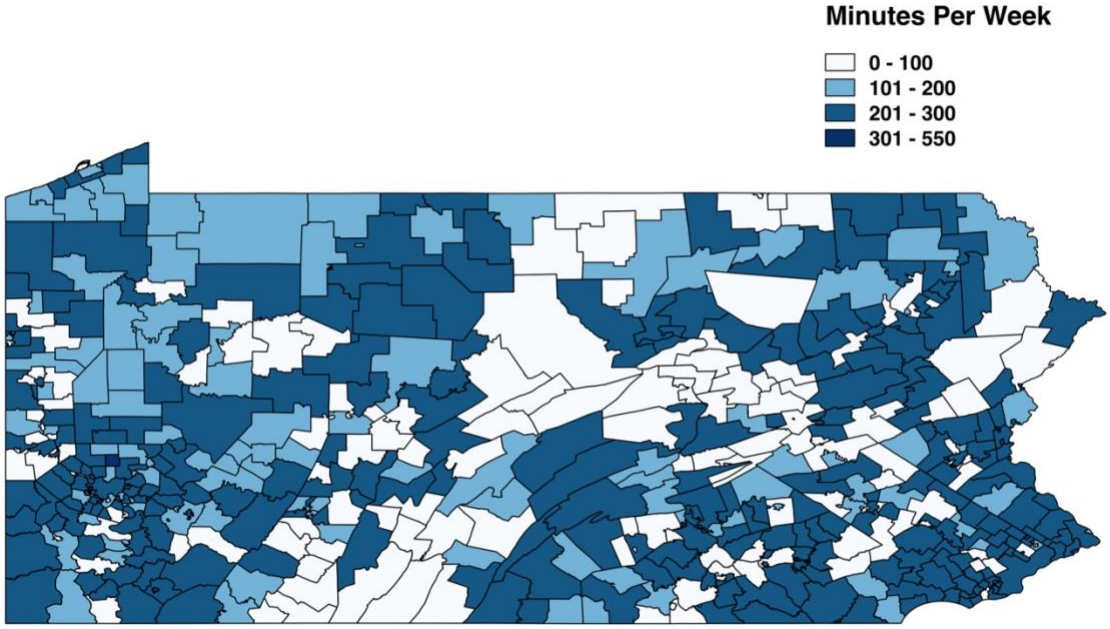


Table 7 presents the probit results for whether a district sets a maximum class size in each of the three grades for which I collected data, as well as the spatial autoregression

results (GS2SLS) considering that the Moran test revealed a lack of spatial independence ($\chi^2 = 5.36$ for 4th grade, 4.88 for 8th grade, and 5.96 for high school, $p > 0.05$). I included only the lagged dependent variable, as the error term was not significant. In Table 7, there are few significant predictors, save for the finding that suburban districts are significantly more likely to provide mandatory teacher preparation periods, and urban districts are significantly less likely. This finding suggests that districts with higher capacity (suburban districts, on average) may be more willing to ensure teachers have protected time and thus include this provision in the contract. Furthermore, the significant Moran test results suggests that districts are willing to provide the same prep time guarantees as their neighbors.

Table 7. Probit Models and Spatial Autoregression (GS2SLS) Results for Mandatory Teacher Preparation Time

	Prep Time 4th?	Prep Time 8th?	Prep Time High?	Minutes (4th)	Minutes (8th)	Minutes (High)
PPE ($1000s)	-0.009	-0.007	-0.010	-1.507	-1.567	-2.118
	(0.006)	(0.006)	(0.006)	(1.571)	(1.617)	(1.633)
Enroll (1000s)	0.013	0.015	0.016	0.052	0.531	0.334
	(0.012)	(0.012)	(0.012)	(0.787)	(0.815)	(0.828)
% Low-Inc.	-0.000	-0.000	-0.001	-0.074	0.289	0.198
	(0.002)	(0.002)	(0.002)	(0.429)	(0.442)	(0.446)
% IEP	0.002	0.003	0.005	-0.807	-0.420	0.091
	(0.006)	(0.006)	(0.006)	(1.408)	(1.452)	(1.468)
% ELL	-0.001	0.002	-0.000	2.730	1.875	1.574
	(0.011)	(0.011)	(0.011)	(2.171)	(2.240)	(2.268)
S-T Ratio	-0.007	-0.009	-0.010	-1.162	-1.902	-3.172
	(0.014)	(0.014)	(0.014)	(3.313)	(3.410)	(3.444)
% Local Tax	0.000	0.000	-0.000	0.048	0.412	0.440
	(0.002)	(0.002)	(0.002)	(0.386)	(0.398)	(0.402)
Rep. Board	-0.045	-0.055	-0.047			
	(0.053)	(0.054)	(0.054)			
Suburban	0.128**	0.109*	0.104*	32.459*	20.856	22.030

	(0.050)	(0.051)	(0.051)	(12.630)	(13.003)	(13.136)
Town	-0.024	-0.023	-0.035	-4.757	-6.748	-8.111
	(0.055)	(0.056)	(0.057)	(13.499)	(13.878)	(14.018)
Urban	-0.555	-0.646*	-0.647*	-4.221	-26.146	-23.311
	(0.334)	(0.258)	(0.260)	(29.329)	(30.199)	(30.508)
% Repub.				-0.474	-0.702	-0.717
				(0.401)	(0.413)	(0.417)
N	499	499	499	499	499	499
pseudo R^2	0.053	0.050	0.049	0.073	0.061	0.056

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$

Transfers and Vacancies

The final set of contract items, before I turn to the overall measures of strength, classroom, and teacher qualification indicators, considers the extent of seniority-based hiring and transfer decisions. Figures 24-27 illustrate the relative rarity of seniority hiring compared to the commonplace practice of simply posting vacancies, which occurs both inside school buildings and online. These are also spatially dependent ($\chi^2 = 6.80$ for seniority, 23.47 for involuntary, and 8.82 for fair share fees), while existing employees filling vacancies ($\chi^2 = 0.21$) and the posting of vacancies ($\chi^2 = 3.68$) are spatially independent, most likely because they are either quite common (posting positions) or rare (seniority hiring first). Neighboring districts are more likely to match each other's seniority-based protections and whether to require a fair share fee.

Figure 24. Districts Where Teachers with Seniority Fill New Positions

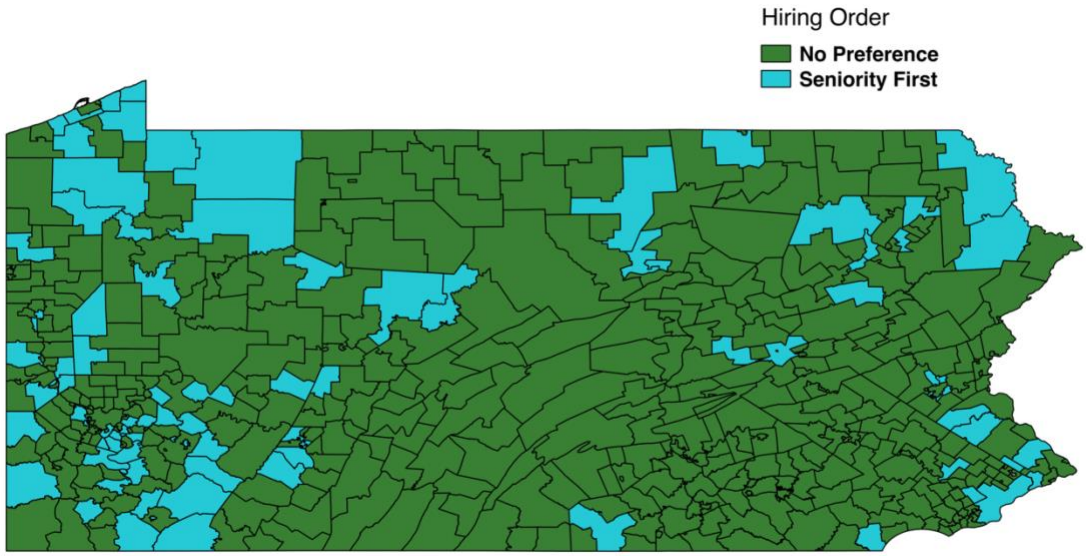


Figure 25. Districts Where Seniority Decides Involuntary Transfers

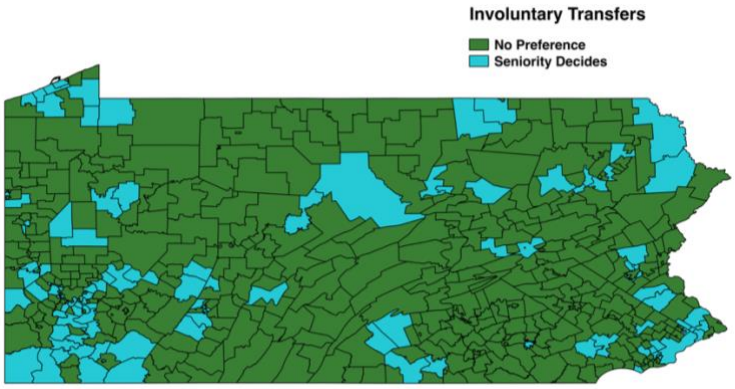


Figure 26. Districts Where Existing Employees Fill Vacancies

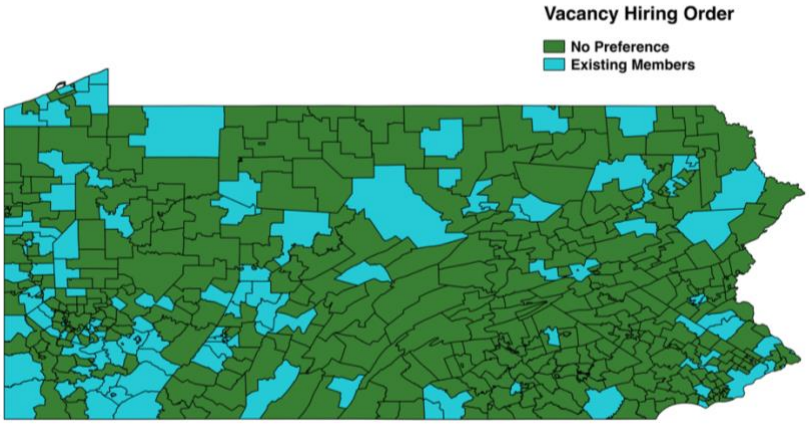


Figure 27. Districts Where All Vacancies Must Be Posted

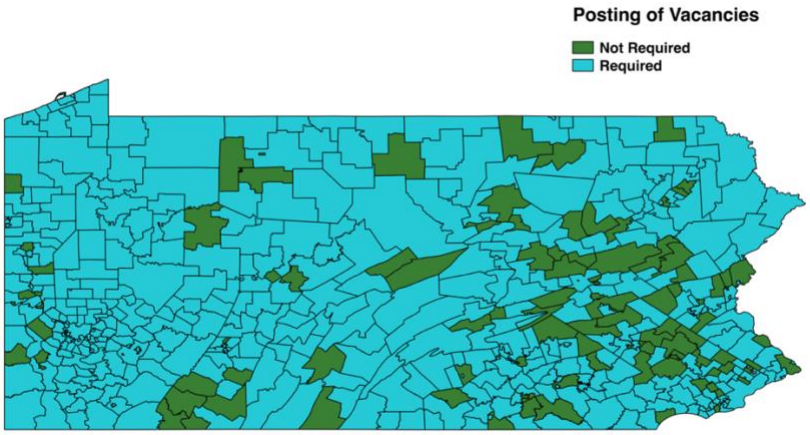


Table 8 presents the probit models for the above items. All of the items, with the exception of whether vacancies are required to be posted, are significantly less likely to occur in districts governed by Republican school boards. These findings match the existing literature (Finger 2017; Moe, 2017) on what we know about Republican and conservative preferences against seniority-based hiring and in favor of performance measures such as value-added scores (Tetlock et al., 2013). The fact that these relationships do not occur among several of the previously studied items provides a strong foundation for understanding the contract issues best related to political party affiliation, either within the electorate or within the school board. The fair share finding also matches the preferences of the Republican plantiff in the *Janus* case.

Table 8. Probit Models for Hiring, Vacancies, and Fair Share Fees

	Senior Vacancies?	Senior Transfer?	Vacancy Order?	Vacancy Posted?	Fair Share Fee?
PPE ($1000s)	0.005	0.005	0.007	0.006	0.007
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Enroll (1000s)	0.007	0.016	0.005	0.022*	0.014
	(0.007)	(0.009)	(0.005)	(0.010)	(0.011)
% Low-Inc.	0.000	0.001	-0.002	0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
% IEP	0.006	0.015*	0.007	-0.004	0.008
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
% ELL	-0.018	-0.015	-0.021	-0.029**	0.007
	(0.011)	(0.010)	(0.012)	(0.009)	(0.011)
S-T Ratio	0.033**	0.035*	0.028	0.023	0.010
	(0.013)	(0.014)	(0.015)	(0.014)	(0.014)
% Local Tax	-0.001	-0.001	-0.003	-0.003	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Rep. Board	-0.107*	-0.111*	-0.210**	-0.028	-0.123*
	(0.047)	(0.051)	(0.054)	(0.051)	(0.052)
Suburban	-0.019	0.033	-0.006	0.021	0.033
	(0.046)	(0.051)	(0.053)	(0.049)	(0.048)
Town	-0.031	-0.072	-0.017	0.022	-0.047
	(0.047)	(0.051)	(0.058)	(0.052)	(0.056)
Urban	-0.052	-0.058	0.128	0.144*	0.001

	(0.097)	(0.108)	(0.161)	(0.057)	(0.130)
N	499	499	499	499	499
pseudo R^2	0.053	0.079	0.062	0.046	0.048

Marginal effects; Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$

Strength, Classroom, and Teacher Qualification Constructs

Unlike previously studied states (Strunk et al., 2018), Pennsylvania demonstrates far less variation because of the requirements of the School Code. I therefore test to see whether Cronbach's α might be a sufficient measure of how strong a contract is based on the findings in the previous section, which respects Pennsylvania's particular policy context and focuses on items for which there is substantial variation among districts. By focusing on these items, I am collapsing some of the gateway provisions into a single data point. Developing a reliable scale with fewer data points represents a meaningful contribution to the extant literature.

By including every single item from Table 1 into a single scale, the α is already somewhat reliable at 0.68. If I remove the extremely rare items and those with weak relationships to district characteristics (bonuses for special education teachers, rows, columns, mandatory prep minutes, school board mediation, and arbitration measures), I can increase the α to 0.82, a more acceptable range. I use this construct throughout the rest of the analysis. When I refer to "strength", I refer to the composite of these 25 items (which include longevity and NBC bonuses, negotiated classes sizes in all three grades, maximum class sizes in all three grades, all items related to teacher evaluation, all items related to seniority, whether a district has mandatory prep periods, and all items related to the school day in Table 1).

I also created classroom and teacher qualification indicators, both of which have an α 0.78 (classroom) or 0.79 (teacher qualification). The classroom construct of 12 items includes only the class size items and the negative consequences for teacher

evaluation. The teacher qualification construct of 15 items includes longevity bonuses, NBC, professional growth plans, preparation time, and all hiring and firing provisions. The three constructs – strength, classroom, and teacher qualification – are positively correlated and significantly with each other, but these relationships are relatively is relatively weak (0.213, 0.147) except for the correlation between strength and teacher qualification (0.59). Consequently, I do not include them simultaneously in any models. Figures 28, 29, and 30 depict this variation across the state. Considering the relative lack of variation in districts with class size maximums, there is a relative rarity of items directly addressing learning environment issues in Figure 29.

Figure 28. CBA Strength

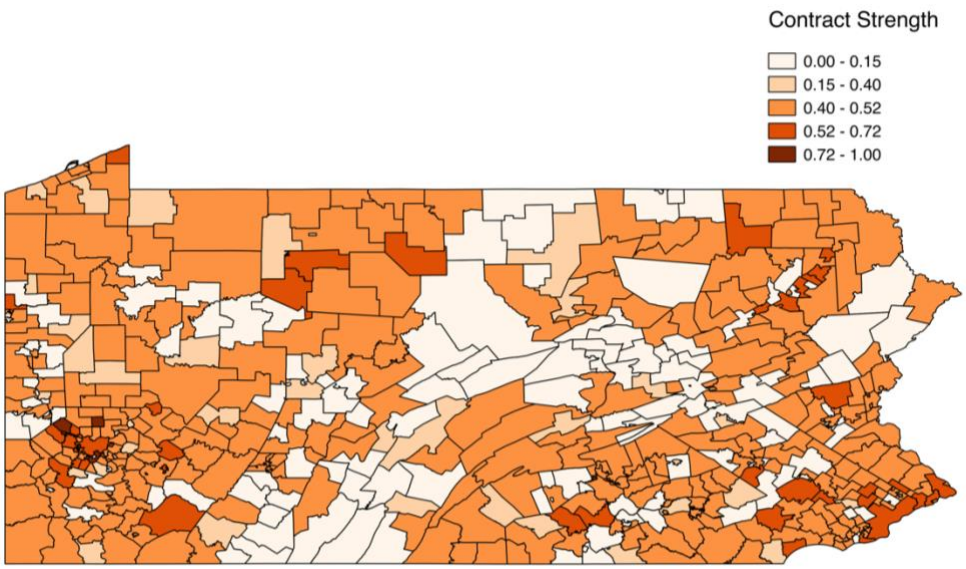


Figure 29. CBA Classroom Items

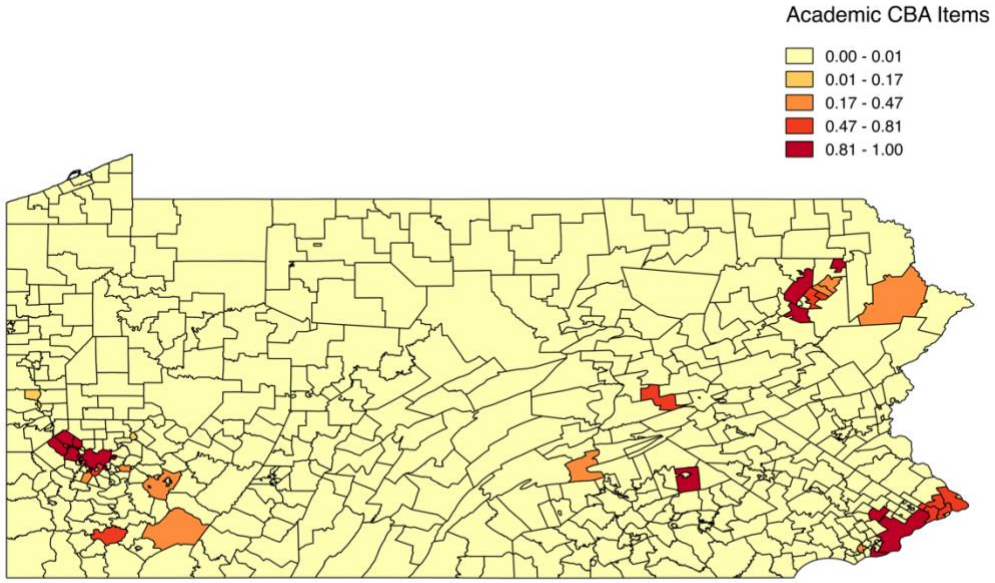
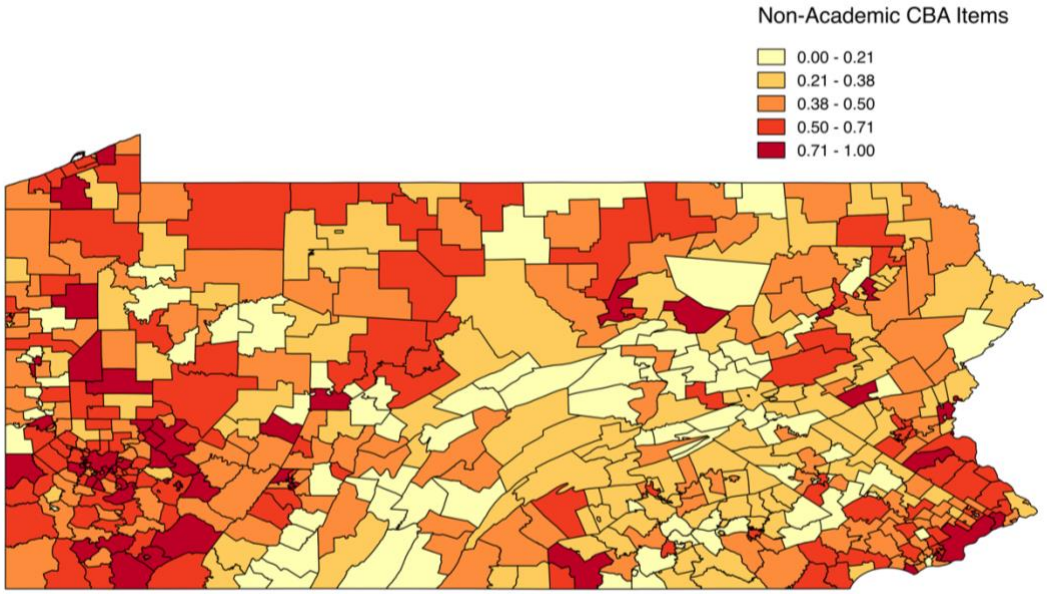


Figure 30. CBA Teacher Qualification Items.



To predict significant relationships for these three constructs, I ran OLS regression models with a Moran test, which was significant ($\chi^2 = 7.33$ for strength,

53.23 for classroom indicators, and 21.18 for teacher qualification indicators, $p > 0.00$). This result is not surprising considering how many of the items in each construct are not spatially independent. In and of itself, this lack of spatial independence is an important finding, suggesting that districts are aware of not only high visibility benefits such as salaries but also lower visibility protections such as seniority rights, which may influence where teachers decide to work (when choice is available).

I use spatial autoregression and present the comparative results in Table 9. I compare the results for both registered voters and school board majorities; they return similar results and patterns of significance. A significant negative relationship between Republican voters and contract strength, classroom, and teacher qualification appears in the OLS models, but it becomes nonsignificant in the GS2SLS models when accounting for spatial dependence with neighboring districts. Not surprisingly, the size (enrollment) of a district also plays a role in whether it agrees to classroom contract items, such as class size caps. Thus, higher enrollment districts are more likely to specify classroom conditions as they grapple with larger class sizes, while Republican districts are significantly less likely to agree to more teacher qualification benefits. These results confirm the hypothesis that Republicans prefer not to concede contract concessions on items that they perceive might be only improving teacher working conditions instead of more directly improving student achievement (Tetlock et al., 2013).

In the next chapter, I turn to the question of student achievement using several models in order to obtain least biased estimates.

Table 9. OLS and Spatial Autoregression (GS2SLS) Results for Strength, Classroom, and Teacher Qualification Items

	Strength (OLS)	Classroom (OLS)	Teacher Qual. (OLS)	Strength (GS2SLS)	Classroom (GS2SLS)	Teacher Qual. (GS2SLS)
PPE ($1000s)	-0.004	-0.001	0.001	-0.004	-0.001	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Enroll (1000s)	0.003	0.007**	0.005**	0.001	0.004**	0.002
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
% Low-Inc.	0.000	0.001	-0.002	0.000	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% IEP	0.000	0.002	0.003	-0.000	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
% ELL	0.002	-0.007	-0.001	0.004	-0.005	0.002
	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)
S-T Ratio	-0.004	-0.006	-0.003	-0.005	-0.003	-0.005
	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
% Local Tax	0.001	0.002*	-0.002	0.001	0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
% Republican	-0.002*	-0.003**	-0.006**	-0.002	-0.001	-0.005**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Suburban	0.052	0.006	0.039	0.053*	0.001	0.047
	(0.027)	(0.023)	(0.027)	(0.027)	(0.021)	(0.026)
Town	-0.031	-0.012	-0.016	-0.014	0.003	0.000
	(0.028)	(0.024)	(0.028)	(0.028)	(0.022)	(0.028)
Urban	-0.032	-0.007	-0.027	-0.036	0.017	-0.029
	(0.063)	(0.054)	(0.063)	(0.062)	(0.049)	(0.061)
_cons	0.472**	0.093	0.818**	0.409*	0.003	0.686**

	(0.168)	(0.144)	(0.168)	(0.166)	(0.131)	(0.164)
N	499	499	499	499	499	499
R^2	0.080	0.136	0.163			
pseudo R^2				0.080	0.166	0.167

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$

CHAPTER 5: STUDENT ACHIEVEMENT AND COLLECTIVE BARGAINING AGREEMENTS

I next analyze relationships among collective bargaining provisions and student achievement. I present this chapter with large caveats, as I only have a single year data and am thus limited in my ability to make causal inferences. However, I did examine the changes across those contracts that changed in 2017-2018 and found, consistent with the limited literature in other states, that these contracts are remarkably consistent over time (Ingle & Wisman, 2018). Those changes that I did notice effective in the 2017-2018 school year were relatively minor – adding or decreasing the number of steps awarded for graduate credentials, for example. Gettysburg Area School District did eliminate a teacher-led merit pay exploratory committee, but it had never adopted merit pay policies in the first place. Interestingly, those districts that did change their pay structures in 2017-2018 opted to increase the incentives for earning additional graduate credits despite a lack of empirical evidence demonstrating a relationship between teacher credentials and teacher quality.

Considering this remarkable sameness over time, I begin with correlational analysis and then employ quasi-experimental techniques to address some of these limitations.

OLS and GS2SLS Analyses

First, I use ordinary least squares (OLS) and spatial autoregression (GS2SLS) to examine relationships before using a machine-learning propensity score generalized boosted model (GBM) approach. I use the adjusted salary throughout in order to account

for local labor market differences. I considered several interaction variables (such as interacting town and urban with the other significant variables), but none of them were significant on their own nor improved the fit of the model. Achievement scores also are not spatially independent – student performance is predicted by the performance of neighboring districts, though we might imagine some important exceptions to this state-level finding (i.e., Philadelphia compared to neighboring wealthy suburbs). The dependent variable of interest is the percentage of students scoring proficient or advanced for that district on the PSSA (grades 3-8). In the Appendix, I also provide the results of the OLS models weighted by enrollment.

I present the results of both the OLS and the GS2SLS models. Table 10 provides the results for the percentage of math students scoring proficient during the 2015-2016 school year and uses three separate predictors of interest based on the CBA coding: strength, classroom, and teacher qualification indicators. Across models and using both OLS and GS2SLS, none of the CBA variables are significant predictors. Other significant negative predictors are consistent with the literature: percentage of low-income students, students on IEPs, and ELLs. Across all models, a higher student-teacher ratio is a significant negative predictor; an increase in the ratio predicts a 1% drop in student achievement proficiency. Two financial factors – percentage of local property tax, and teacher salaries – are significant predictors, whereas per-pupil-expenditures are not significant. Finally, having a Republican school board was a significant positive predictor of math student achievement, predicting a more than 2% increase in passing rates. But this variable is not significant in the ELA nor the math models weighted by enrollment,

and it remains nonsignificant even when removing the two largest cities, Philadelphia and Pittsburgh (heavily Democratic) from the analysis. In contrast, the patterns of significance hold for all of the other variables in the models, weighted by enrollment. The Republican finding, therefore, is not particularly robust.

Table 10. OLS and Spatial Autoregression (GS2SLS) Results for Math Achievement

	OLS			GS2SLS		
	Math % Proficient	Math % Proficient	Math % Proficient	Math % Proficient	Math % Proficient	Math % Proficient
Strength	1.359			1.329		
	(1.764)			(1.723)		
Classroom		-0.170			0.550	
		(2.064)			(2.071)	
Teacher Qual.			3.205			2.783
			(1.726)			(1.716)
PPE ($1000s)	-0.064	-0.071	-0.072	-0.120	-0.124	-0.124
	(0.139)	(0.139)	(0.138)	(0.136)	(0.136)	(0.135)
Enroll (1000s)	-0.029	-0.024	-0.044	-0.033	-0.033	-0.043
	(0.063)	(0.065)	(0.063)	(0.061)	(0.063)	(0.062)
% Low-Inc.	-0.470**	-0.469**	-0.467**	-0.450**	-0.450**	-0.448**
	(0.035)	(0.035)	(0.035)	(0.036)	(0.036)	(0.035)
% IEP	-0.752**	-0.750**	-0.765**	-0.703**	-0.705**	-0.718**
	(0.118)	(0.118)	(0.118)	(0.117)	(0.118)	(0.118)
% ELL	-0.454*	-0.451*	-0.451*	-0.523**	-0.515**	-0.522**
	(0.182)	(0.183)	(0.181)	(0.179)	(0.179)	(0.178)
S-T Ratio	-0.922**	-0.932**	-0.915**	-0.930**	-0.932**	-0.924**
	(0.285)	(0.285)	(0.284)	(0.277)	(0.278)	(0.277)
% Local Tax	0.080*	0.081*	0.087**	0.098**	0.099**	0.103**

	(0.032)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
Rep. Board	2.206*	2.129*	2.444*	2.331*	2.310*	2.487*
	(1.037)	(1.047)	(1.043)	(1.033)	(1.038)	(1.036)
Suburban	-2.543	-2.417	-2.664	-3.040	-2.940	-3.109
	(2.298)	(2.294)	(2.289)	(2.199)	(2.196)	(2.195)
Town	-1.386	-1.391	-1.329	-2.326	-2.292	-2.271
	(2.481)	(2.483)	(2.474)	(2.403)	(2.404)	(2.399)
Rural	-4.094	-4.052	-4.044	-4.598	-4.554	-4.551
	(2.438)	(2.438)	(2.430)	(2.360)	(2.360)	(2.355)
Sal. ($1000s)	0.123*	0.127*	0.114*	0.119*	0.120*	0.112*
	(0.050)	(0.050)	(0.050)	(0.052)	(0.052)	(0.052)
Years of Exp.	0.237	0.226	0.281	0.231	0.226	0.261
	(0.228)	(0.228)	(0.228)	(0.220)	(0.220)	(0.221)
Constant	78.574**	79.061**	77.334**	77.869**	78.348**	77.041**
	(7.017)	(6.993)	(7.030)	(6.943)	(6.923)	(6.949)
N	499	499	499	499	499	499
R^2	0.688	0.687	0.689	0.687	0.687	0.689

Standard errors in parentheses. Urban is the reference group.

* $p < 0.05$, ** $p < 0.01$

Table 11 presents the results for ELA, where there are some important differences compared to Table 10. Again, the CBA constructs are not significant predictors, and the percentage of low-income students, students on IEPs, LEP students, and higher student-teacher ratios are significant negative predictors for ELA student achievement. Teacher salaries are no longer significant, but percentage of local property taxes remains significant and positive across models. Unlike math achievement, for ELA achievement, years of experience becomes statistically significant. Increasing the average years of experience in a district by one year increases the percentage of students passing by more than one half of one percent.

Years of experience, however, is not significant in the models weighted by enrollment (see Appendix A). And while the Republican school board predictor is still positive in the enrollment models, it is no longer significant in every OLS model, though it is significant when controlling for spatial interdependence (trends in neighboring districts). Nearly 2% more students in Republican-controlled districts pass the mathematics exam in grades 3-8. These results provide slightly suggestive evidence of Republican emphasis on mathematics over English language arts, mirroring a national trend where conservatives de-emphasize liberal arts in favor of science and math (Cohen, 2016). The experience finding may also speaks to the important of teacher experience in ELA, for which preparing students for state tests of reading comprehension may require more experience and content knowledge (Jennings & Bearak, 2014). Though this

experience finding is not significant when weighting by enrollment, it is still double the size of the math coefficients across all models.

Table 11. OLS and Spatial Autoregression (GS2SLS) Results for ELA Achievement

	OLS			GS2SLS		
	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient
Strength	1.832			1.926		
	(1.437)			(1.400)		
Classroom		0.985			0.942	
		(1.682)			(1.689)	
Teacher Qual.			2.228			1.840
			(1.408)			(1.400)
PPE ($1000s)	-0.096	-0.103	-0.106	-0.104	-0.109	-0.111
	(0.113)	(0.113)	(0.113)	(0.110)	(0.110)	(0.110)
Enroll (1000s)	-0.053	-0.055	-0.061	-0.066	-0.068	-0.070
	(0.051)	(0.053)	(0.052)	(0.050)	(0.051)	(0.050)
% Low-Inc.	-0.481**	-0.482**	-0.479**	-0.459**	-0.460**	-0.458**
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
% IEP	-0.642**	-0.643**	-0.650**	-0.640**	-0.644**	-0.650**
	(0.096)	(0.096)	(0.096)	(0.096)	(0.096)	(0.096)
% LEP	-0.488**	-0.478**	-0.484**	-0.517**	-0.505**	-0.513**
	(0.148)	(0.149)	(0.148)	(0.145)	(0.146)	(0.145)
S-T Ratio	-0.707**	-0.711**	-0.708**	-0.731**	-0.732**	-0.729**
	(0.232)	(0.232)	(0.232)	(0.225)	(0.226)	(0.225)
% Local Tax	0.069**	0.069**	0.075**	0.086**	0.086**	0.090**
	(0.026)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)

Rep. Board	1.570	1.562	1.694*	1.813*	1.788*	1.868*
	(0.844)	(0.853)	(0.851)	(0.841)	(0.845)	(0.845)
Suburban	-1.197	-1.049	-1.201	-1.355	-1.210	-1.323
	(1.871)	(1.870)	(1.868)	(1.783)	(1.783)	(1.784)
Town	0.238	0.245	0.275	-0.141	-0.081	-0.091
	(2.021)	(2.023)	(2.019)	(1.951)	(1.955)	(1.952)
Rural	-1.818	-1.763	-1.756	-2.019	-1.951	-1.962
	(1.985)	(1.987)	(1.983)	(1.916)	(1.919)	(1.917)
Sal. ($1000s)	0.076	0.079	0.072	0.058	0.059	0.055
	(0.041)	(0.041)	(0.041)	(0.043)	(0.043)	(0.043)
Years of Exp.	0.522**	0.515**	0.546**	0.540**	0.533**	0.553**
	(0.185)	(0.186)	(0.186)	(0.178)	(0.179)	(0.179)
Constant	92.252**	92.927**	91.709**	91.963**	92.668**	91.775**
	(5.715)	(5.699)	(5.737)	(5.649)	(5.638)	(5.666)
N	499	499	499	499	499	499
R^2 / pseudo R^2	0.761	0.760	0.761	0.760	0.760	0.761

Standard errors in parentheses. Urban is the reference group.
* $p < 0.05$, ** $p < 0.01$

When looking at historically disadvantaged students, these relationships change somewhat in Tables 12-13. Pennsylvania defines historically disadvantaged students as a non-duplicative count of students with disabilities, economically disadvantaged students, and English learners enrolled for a full academic year (PDE, 2018). When looking only at historically disadvantaged students, a Republican school board is no longer a significant predictor of student achievement, and teacher salaries are no longer significant in any of the GS2SLS models. It is also not significant in the weighted by enrollment models (Appendix A). Years of teaching experience remains significant for ELA achievement. Interestingly, adjusted salaries become significant and positive in the weighted enrollment models. None of the patterns of significance for the other district characteristics change.

In Tables 14-15, I present the results with one change – using the percentage of registered Republicans instead of a Republican school board. Here the significant positive relationship between percentage of Republicans and achievement in both subjects re-emerges, but only when controlling for spatial dependence. It is not significant when using OLS weighted by enrollment.

In Table 16, I present the results for high school scores using the force option, as some districts do not have high school students and thus cannot be included for their spillover effects. The student-teacher ratio relationship is larger for historically disadvantaged students in biology and literature, while it becomes nonsignificant in algebra. The relationship between Republican school boards, mirroring the math achievement results in grades 3-8, is strongest when looking at algebra achievement for

all students. Teacher salaries predictors are significant and positive in math and biology only. In the Appendix, I provide the weighted enrollment results using registered Republicans (Table 24); the Republican school board variable in those models are not significant. In the weighted by enrollment models, having more registered Republicans in a district predicts significantly higher achievement in math and biology, consistent with the GS2SLS results.

I interpret the results of these correlational analyses as follows by highlighting these findings. First, there is a significant positive relationship between Republican school boards and math achievement when looking at all students, but this relationship is weaker or nonsignificant when looking at the historically disadvantaged super subgroup, or when looking at ELA. This relationship recurs when looking at registered Republicans and high school math and biology achievement. Second, teacher salaries are significant positive predictors only when looking at math achievement for all students (with the except of a smaller, significant relationship between salaries and ELA achievement in the enrollment weighted models). These are only correlations, though it is important to note that the percentage of Republicans and Republican school boards are significantly and negatively correlated with per pupil expenditures (Pearson's r = -0.1137 to -0.1414).

Third, there is only one model out of dozens with any significant contract predictors for achievement (a positive relationship between teacher qualification items for historically disadvantaged math students in Table 14). Fourth, years of experience is a significant positive predictor in ELA for both all students and historically disadvantaged students. Fifth, a higher student-teacher ratio predicts (moving from 10 to 1 to 11 to 1, for

example), predicts a 0.71-1.499 drop in student proficiency, and this relationship is larger in magnitude for historically disadvantaged students. Though correlational, these findings are important in establishing the significance of student-teacher ratios and salaries (important) compared to contentious contractual issues such as seniority hiring and firing (less important) when looking at student achievement across school districts in Pennsylvania. I also demonstrate the importance of controlling for the achievement in neighboring districts when testing significant predictors within an a priori theoretical framework.

In the next stage of analysis, propensity score weighting, I do not expect to find significant quasi-experimental effects of contract items, consistent with the results of the GS2SLS models.

Table 12. OLS and GS2SLS Results for Math Historically Disadvantaged Using Republican School Boards

	% Math Proficient	% Math Proficient	% Math Proficient	% Math Proficient	% Math Proficient	% Math Proficient
Strength	1.647			1.702		
	(1.639)			(1.602)		
Classroom		0.417			0.974	
		(1.918)			(1.926)	
Teacher Qual.			3.118			2.594
			(1.604)			(1.596)
PPE ($1000s)	-0.145	-0.153	-0.155	-0.194	-0.200	-0.200
	(0.129)	(0.129)	(0.129)	(0.126)	(0.126)	(0.126)
Enroll (1000s)	-0.005	-0.003	-0.019	-0.016	-0.018	-0.024
	(0.058)	(0.060)	(0.059)	(0.057)	(0.059)	(0.057)
% Low-Inc.	-0.146**	-0.146**	-0.143**	-0.132**	-0.132**	-0.130**
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
% IEP	-0.644**	-0.643**	-0.657**	-0.578**	-0.582**	-0.594**
	(0.109)	(0.110)	(0.109)	(0.109)	(0.110)	(0.109)
% LEP	-0.535**	-0.528**	-0.532**	-0.574**	-0.563**	-0.571**
	(0.169)	(0.170)	(0.169)	(0.166)	(0.167)	(0.166)
S-T Ratio	-0.948**	-0.956**	-0.944**	-0.964**	-0.965**	-0.960**
	(0.265)	(0.265)	(0.264)	(0.258)	(0.258)	(0.258)
% Local Tax	0.042	0.043	0.050	0.061*	0.062*	0.066*
	(0.030)	(0.030)	(0.030)	(0.031)	(0.031)	(0.031)

Rep. Board	1.529	1.486	1.746	1.789	1.778	1.908*
	(0.963)	(0.973)	(0.969)	(0.960)	(0.965)	(0.964)
Suburban	-1.483	-1.341	-1.571	-1.908	-1.782	-1.929
	(2.135)	(2.132)	(2.127)	(2.046)	(2.044)	(2.044)
Town	-0.480	-0.479	-0.426	-1.051	-1.004	-0.990
	(2.305)	(2.308)	(2.299)	(2.235)	(2.237)	(2.233)
Rural	-2.401	-2.351	-2.343	-2.720	-2.662	-2.661
	(2.265)	(2.267)	(2.258)	(2.195)	(2.196)	(2.192)
Sal. ($1000s)	0.099*	0.103*	0.091	0.088	0.089	0.082
	(0.046)	(0.046)	(0.047)	(0.048)	(0.049)	(0.048)
Years of Exp.	0.253	0.243	0.293	0.250	0.244	0.275
	(0.211)	(0.212)	(0.212)	(0.205)	(0.205)	(0.205)
Constant	52.593**	53.193**	51.503**	52.048**	52.674**	51.432**
	(6.520)	(6.500)	(6.532)	(6.455)	(6.438)	(6.464)
N	499	499	499	499	499	499
R^2	0.372	0.371	0.375			
pseudo R^2				0.370	0.369	0.374

Table 13. OLS and GS2SLS Results for ELA Historically Disadvantaged Using Republican School Boards

	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient
Strength	1.748			1.906		
	(1.555)			(1.516)		
Classroom		0.841			0.640	
		(1.820)			(1.827)	
Teacher Qual.			2.288			1.839
			(1.524)			(1.515)
PPE ($1000s)	-0.155	-0.162	-0.164	-0.176	-0.182	-0.183
	(0.123)	(0.123)	(0.122)	(0.119)	(0.119)	(0.119)
Enroll (1000s)	-0.011	-0.011	-0.019	-0.032	-0.031	-0.035
	(0.055)	(0.057)	(0.056)	(0.054)	(0.055)	(0.054)
% Low-Inc.	-0.178**	-0.178**	-0.175**	-0.157**	-0.157**	-0.156**
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
% IEP	-0.677**	-0.678**	-0.686**	-0.653**	-0.655**	-0.663**
	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)
% LEP	-0.523**	-0.514**	-0.519**	-0.535**	-0.525**	-0.532**
	(0.160)	(0.161)	(0.160)	(0.157)	(0.158)	(0.157)
S-T Ratio	-0.848**	-0.853**	-0.848**	-0.896**	-0.899**	-0.894**
	(0.251)	(0.251)	(0.251)	(0.244)	(0.244)	(0.244)

% Local Tax	0.033	0.034	0.040	0.055	0.056	0.059*
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Rep. Board	0.897	0.882	1.031	1.344	1.299	1.393
	(0.914)	(0.923)	(0.921)	(0.910)	(0.915)	(0.915)
Suburban	0.175	0.318	0.159	-0.112	0.036	-0.077
	(2.025)	(2.023)	(2.021)	(1.933)	(1.933)	(1.934)
Town	0.949	0.955	0.987	0.565	0.619	0.615
	(2.186)	(2.189)	(2.184)	(2.114)	(2.118)	(2.115)
Rural	-0.922	-0.869	-0.862	-1.101	-1.036	-1.043
	(2.148)	(2.150)	(2.145)	(2.075)	(2.079)	(2.076)
Sal. ($1000s)	0.099*	0.101*	0.094*	0.083	0.085	0.080
	(0.044)	(0.044)	(0.044)	(0.046)	(0.046)	(0.046)
Years of Exp.	0.524**	0.517*	0.550**	0.527**	0.518**	0.540**
	(0.201)	(0.201)	(0.202)	(0.193)	(0.194)	(0.194)
Constant	67.056**	67.699**	66.451**	66.685**	67.369**	66.488**
	(6.184)	(6.165)	(6.207)	(6.115)	(6.102)	(6.133)
N	499	499	499	499	499	499
R^2	0.452	0.450	0.453			
pseudo R^2				0.450	0.449	0.451

Table 14. OLS and GS2SLS Results for Math Historically Disadvantaged Using Registered Republicans

	% Math Proficient	% Math Proficient	% Math Proficient	% Math Proficient	% Math Proficient	% Math Proficient
Strength	1.689			1.706		
	(1.639)			(1.600)		
Classroom		0.471			1.069	
		(1.916)			(1.927)	
Teacher Qual.			3.658*			3.050
			(1.641)			(1.621)
PPE ($1000s)	-0.135	-0.143	-0.140	-0.177	-0.182	-0.178
	(0.130)	(0.130)	(0.129)	(0.127)	(0.127)	(0.126)
Enroll (1000s)	0.008	0.009	-0.006	-0.001	-0.004	-0.009
	(0.058)	(0.060)	(0.059)	(0.057)	(0.058)	(0.057)
% Low-Inc.	-0.142**	-0.142**	-0.136**	-0.127**	-0.127**	-0.123**
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
% IEP	-0.633**	-0.633**	-0.644**	-0.568**	-0.572**	-0.585**
	(0.110)	(0.110)	(0.109)	(0.109)	(0.110)	(0.109)
% LEP	-0.510**	-0.503**	-0.501**	-0.544**	-0.532**	-0.538**
	(0.169)	(0.169)	(0.168)	(0.166)	(0.167)	(0.166)
S-T Ratio	-0.935**	-0.943**	-0.922**	-0.947**	-0.947**	-0.936**
	(0.265)	(0.265)	(0.264)	(0.258)	(0.258)	(0.258)
% Local Tax	0.040	0.041	0.048	0.060	0.060*	0.065*
	(0.030)	(0.030)	(0.030)	(0.031)	(0.031)	(0.031)

% Rep	0.057	0.056	0.074*	0.071*	0.071*	0.082*
	(0.032)	(0.032)	(0.033)	(0.034)	(0.034)	(0.035)
Suburban	-1.515	-1.370	-1.647	-1.967	-1.843	-2.024
	(2.134)	(2.131)	(2.124)	(2.045)	(2.042)	(2.042)
Town	-0.792	-0.782	-0.858	-1.410	-1.366	-1.431
	(2.321)	(2.324)	(2.312)	(2.251)	(2.253)	(2.248)
Rural	-2.757	-2.696	-2.839	-3.132	-3.077	-3.169
	(2.286)	(2.287)	(2.277)	(2.217)	(2.218)	(2.213)
Sal. ($1000s)	0.106*	0.109*	0.097*	0.093	0.094	0.087
	(0.047)	(0.047)	(0.047)	(0.049)	(0.049)	(0.048)
Years of Exp.	0.221	0.212	0.258	0.214	0.208	0.237
	(0.213)	(0.213)	(0.213)	(0.206)	(0.206)	(0.206)
Constant	50.535**	51.205**	48.362**	49.497**	50.108**	48.099**
	(6.767)	(6.741)	(6.829)	(6.734)	(6.713)	(6.779)
N	499	499	499	499	499	499
R^2	0.373	0.371	0.378			
pseudo R^2				0.371	0.370	0.376

Table 15. OLS and GS2SLS Results for ELA Historically Disadvantaged Using Registered Republicans

	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient	% ELA Proficient
Strength	1.782			1.942		
	(1.555)			(1.512)		
Classroom		0.891			0.790	
		(1.818)			(1.827)	
Teacher Qual.			2.649			2.272
			(1.560)			(1.536)
PPE ($1000s)	-0.148	-0.155	-0.154	-0.157	-0.162	-0.161
	(0.123)	(0.123)	(0.123)	(0.120)	(0.120)	(0.120)
Enroll (1000s)	-0.003	-0.004	-0.011	-0.021	-0.021	-0.025
	(0.055)	(0.057)	(0.056)	(0.054)	(0.055)	(0.054)
% Low-Inc.	-0.175**	-0.176**	-0.170**	-0.150**	-0.150**	-0.147**
	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)	(0.032)
% IEP	-0.670**	-0.671**	-0.678**	-0.642**	-0.645**	-0.652**
	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)
% LEP	-0.508**	-0.499**	-0.501**	-0.510**	-0.499**	-0.506**
	(0.160)	(0.161)	(0.160)	(0.157)	(0.158)	(0.157)
S-T Ratio	-0.839**	-0.844**	-0.833**	-0.877**	-0.879**	-0.870**
	(0.251)	(0.252)	(0.251)	(0.244)	(0.244)	(0.244)

% Local Tax	0.032	0.033	0.039	0.054	0.055	0.059*
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
% Rep	0.036	0.035	0.047	0.066*	0.065*	0.073*
	(0.030)	(0.030)	(0.031)	(0.032)	(0.033)	(0.033)
Suburban	0.155	0.300	0.110	-0.181	-0.031	-0.179
	(2.024)	(2.022)	(2.020)	(1.928)	(1.928)	(1.928)
Town	0.743	0.754	0.699	0.166	0.228	0.158
	(2.202)	(2.205)	(2.199)	(2.126)	(2.129)	(2.126)
Rural	-1.156	-1.097	-1.193	-1.552	-1.480	-1.560
	(2.169)	(2.170)	(2.165)	(2.093)	(2.096)	(2.093)
Sal. ($1000s)	0.103*	0.105*	0.098*	0.088	0.089	0.084
	(0.044)	(0.044)	(0.044)	(0.046)	(0.046)	(0.046)
Years of Exp.	0.504*	0.497*	0.526**	0.489*	0.481*	0.502*
	(0.202)	(0.202)	(0.203)	(0.195)	(0.195)	(0.195)
Constant	65.707**	66.389**	64.364**	63.931**	64.667**	63.172**
	(6.420)	(6.396)	(6.494)	(6.377)	(6.360)	(6.431)
N	499	499	499	499	499	499
R^2	0.452	0.451	0.454			
pseudo R^2				0.449	0.448	0.451

Table 16. GS2SLS Results for High School Student Achievement, All and Historically Disadvantaged (Dis.)

	Algebra (All)	Algebra (Dis.)	Biology (All)	Bio (Dis.)	Literature (All)	Literature (Dis.)
Strength	0.118	-0.879	0.552	-0.465	-0.574	-0.966
	(2.007)	(2.662)	(2.046)	(2.674)	(1.698)	(2.535)
PPE ($1000s)	-0.141	0.054	-0.356*	-0.228	-0.219	-0.188
	(0.158)	(0.209)	(0.161)	(0.210)	(0.134)	(0.199)
Enroll (1000s)	0.041	0.042	0.027	0.046	0.053	0.067
	(0.071)	(0.094)	(0.073)	(0.094)	(0.060)	(0.090)
% Low-Inc.	-0.385**	-0.160**	-0.474**	-0.273**	-0.285**	-0.125*
	(0.042)	(0.055)	(0.042)	(0.054)	(0.035)	(0.052)
% IEP	-0.745**	-0.835**	-0.759**	-0.776**	-0.737**	-0.911**
	(0.139)	(0.185)	(0.142)	(0.184)	(0.117)	(0.176)
% LEP	-0.711**	-0.492	-0.723**	-0.578*	-0.671**	-0.390
	(0.207)	(0.272)	(0.211)	(0.273)	(0.175)	(0.260)
S-T Ratio	-0.708*	-0.481	-1.229**	-1.449**	-0.794**	-1.103**
	(0.325)	(0.437)	(0.332)	(0.439)	(0.275)	(0.417)
% Local Tax	0.065	-0.034	0.052	-0.053	0.080*	-0.002
	(0.038)	(0.050)	(0.039)	(0.050)	(0.032)	(0.048)
Rep. Board	0.147**	0.118*	0.117**	0.090	0.083*	0.007
	(0.043)	(0.054)	(0.043)	(0.053)	(0.035)	(0.051)

Suburban	-2.240	0.763	-2.280	0.555	-0.663	2.806
	(2.560)	(3.390)	(2.617)	(3.432)	(2.180)	(3.249)
Town	0.884	2.078	1.214	2.678	1.677	3.834
	(2.820)	(3.721)	(2.878)	(3.751)	(2.392)	(3.559)
Rural	-1.362	0.596	-2.071	0.194	-0.205	2.361
	(2.772)	(3.657)	(2.829)	(3.686)	(2.351)	(3.497)
Sal. ($1000s)	0.136*	0.133	0.147*	0.168*	0.067	0.077
	(0.060)	(0.078)	(0.061)	(0.077)	(0.050)	(0.074)
Years of Exp.	0.009	-0.131	0.260	-0.125	0.176	0.116
	(0.258)	(0.345)	(0.264)	(0.347)	(0.219)	(0.328)
Constant	97.098**	71.274**	107.957**	91.877**	106.262**	95.722**
	(8.426)	(11.315)	(8.570)	(11.287)	(7.086)	(10.751)
N	496	490	496	491	496	492
pseudo R^2	0.580	0.221	0.625	0.291	0.562	0.192

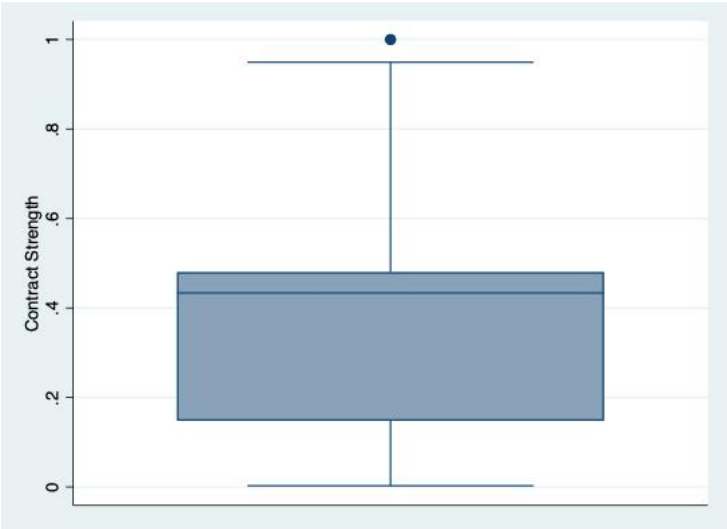
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Propensity Score Weighting

Considering the results in the previous section and my research questions, I next turn to obtaining estimates for contract strength, classroom items, teacher qualification items, high-profile provisions, increased teacher salaries, and decreased student-teacher ratios. I did not expect to find significant results for the CBA constructs considering the absence of significant relationships in the previous section.

In order to create binary treatment variables, I examined different potential cut-off points for strength, for classroom, and for teacher qualification items. For strength, I examined cut-off points for four, three, and two quantiles. These all produced values between 0.40 and 0.50, which I do not consider to be substantively different from each other. Strength also is not normally distributed, as shown in Figure 31 below, and has a leftward skew. Thus, I created two variables using a strongest cut-off above 0.50, which applies to 62 districts. I then created a weakest variable for values below 0.25, which applies to 131 districts.

Figure 31. Distribution of Contract Strength.



For the classroom indicator, there was little variation, and most districts had a value of 0 (400 districts out of 499). Thus, instead of using cut-offs, I generated a classroom binary variable for all of the districts with a value greater than 0. For the teacher qualification indicator, I created a variable for the top-third of districts, all of which had values greater than 0.50.

Table 17 presents the balance of the treatment and comparison groups for contract strength after using the machine learning TWANG package in R (Ridgeway, McCaffrey, Morral, Burgette, & Griffin, 2014). There are no significant differences between the treatment and comparison groups after weighting. I present the results for contract strength greater than 0.50 rather than show every table in this chapter. I provide the full list of balance tables in the Appendix in Tables 25-36.

Table 17. Balance of the Treatment and Comparison Groups, Strength > 0.50

	Before Weighting			After Weighting		
	Treatment	Control	*t*	Treatment	Control	*t*
PPE ($100s)	18.342	17.265	2.305*	18.342	18.286	0.102
	(3.487)	(3.317)		(3.487)	(3.865)	
Enroll (1000s)	5.859	2.781	1.437	5.859	3.647	1.027
	(16.963)	(2.521)		(16.963)	(3.048)	
% Low-Inc	43.768	43.126	0.236	43.768	42.224	0.49
	(20.638)	(16.582)		(20.638)	(19.928)	
% IEP	18.896	18.085	1.31	18.896	18.485	0.598
	(4.708)	(3.634)		(4.708)	(4.051)	
% LEP	2.359	1.224	2.801**	2.359	2.488	-0.227
	(3.071)	(2.495)		(3.071)	(4.083)	
S-T Ratio	14.184	13.987	0.734	14.184	14.179	0.018
	(2.040)	(1.596)		(2.040)	(1.641)	
% Local	58.467	50.697	3.241**	58.467	58.253	0.081
	(17.666)	(18.536)		(17.666)	(18.646)	
% Repub.	37.967	46.725	-4.879**	37.967	40.957	-1.505
	(13.385)	(12.775)		(13.385)	(13.343)	
Salary (1000s)	72.37	69.987	2.141*	72.37	72.112	0.212
	(8.176)	(8.772)		(8.176)	(8.498)	
Yrs of Exp	14.126	14.279	0.510	14.126	14.200	0.766
	(1.719)	(1.747)		(1.719)	(1.735)	

Standard deviations in parentheses, * $p < 0.05$, ** $p < 0.01$

For most of the "treatments," I was able to achieve balance, though not in two cases. Because the percentage of Republicans in a district is significantly and negatively correlated with the classroom construct (Pearson's r = -0.26), the difference between the two groups is still significant (see Table 26). The other significant differences occur when attempting to match on partisanship (either majority Republican school board or registered voters). The outcomes in Table 26, 34, and 35 highlight the need to use doubly robust estimations, so I include all of the covariates in the generalized boosted models (GBM) after weighting. I caution against overinterpretation of either measure of partisanship as a treatment, as it theoretically difficult to imagine shifting a school district dramatically in terms of its partisan composition. Party self-identification often begins in childhood (Campbell, Converse, Miller, & Stokes, 1980).

I also present the results of the PSW procedure for adjusted salaries in Table 18, which was perhaps the most challenging iteration because of the significant positive relationships among teacher salaries, per pupil expenditures, and percentage of local tax revenues (Pearson's r ranges from 0.24-0.35). Overall, the largely successful weighting procedures evidenced in Tables 17 and 18, and in Tables 25-36, highlight the strength of using the TWANG package to achieve optimal balance at a setting of 10,000 iterations, as well as the importance of including all of the covariates in the model in order to make the treatment and comparison districts as similar as possible.

Table 18. Balance of the Treatment and Comparison Groups, Top Salary ($76k)

	Before Weighting			After Weighting		
	Treatment	Control	*t*	Treatment	Control	*t*
PPE ($1000s)	18.995	16.87	5.376**	18.995	18.085	1.935
	(4.088)	(2.892)		(4.088)	(3.063)	
Enroll (1000s)	3.583	3.024	1.213	3.583	3.349	0.533
	(2.994)	(7.258)		(2.994)	(6.429)	
% Low-Inc	37.353	45.141	-4.239**	37.353	39.493	-0.933
	(18.242)	(16.295)		(18.242)	(17.762)	
% IEP	17.089	18.548	-3.816**	17.089	17.635	-1.228
	(3.68)	(3.758)		(3.68)	(3.739)	
% LEP	1.330	1.377	-0.23	1.33	1.300	0.150
	(1.579)	(2.857)		(1.579)	(2.076)	
S-T Ratio	14.18	13.956	1.393	14.18	14.179	0.003
	(1.506)	(1.702)		(1.506)	(1.545)	
% Local	59.61	49.034	5.657**	59.61	56.662	1.335
	(18.116)	(18.011)		(18.116)	(18.469)	
% Repub.	43.499	46.344	-2.254*	43.499	45.078	-1.154
	(11.737)	(13.538)		(11.737)	(11.60)	
Rest. (0-1)	0.386	0.342	2.046*	0.386	0.388	-0.124
	(0.203)	(0.209)		(0.203)	(0.187)	
Yrs of Exp	15.106	13.980	7.092**	15.106	14.772	1.911
	(1.464)	(1.738)		(1.464)	(1.533)	

Standard deviations in parentheses
* $p < 0.05$, ** $p < 0.01$

To summarize, I am obtaining estimates for strongest contracts (above 0.50), weakest (0.25), most classroom items (values above 0), and most teacher qualification items (values above 0.50). I use the voter registration file and the percentage of Registered republican variable in order to make use of the of the most precise measure of a district's political composition, but I also include the Republican school board majority as a variable. As a reminder, voters are typically selecting school directors who have won both the Democratic *and* the Republican primary, which makes the actual registration of voters a more reliable indicator the district's political leanings. I also analyze individual

contract items that I hypothesize might have significant relationships to student achievement based on the literature discussed in Chapter 2. These additional treatment variables are: 1) bonuses for National Board certification, 2) class size maximums, 3) negative consequences for poor teacher evaluations, 4) seniority-based layoffs, 5) mandatory prep time (4th grade), and 6) seniority-based hiring.

For the continuous treatment salary variable, I also examined quartiles and percentiles across the state and determined treatment cut-off points. An adjusted average salary of $70,000 represents the top 241-paying districts in the state, whereas being in the top 25% of districts represents an adjusted average of $76,000. I use this difference of $6,000 as a treatment variable. Swain, Rodriguez, and Springer (2019) find this amount to be sufficient for a positive impact on teacher retention.

There are, however, limitations to the use of an average salary variable, as a higher average salary could represent teachers with more credentials and more experience, not just a higher starting salary table. On the other hand, the average adjusted salary measure incorporates both of these advantages, including National Board certification. Another major caveat, though, is that there is little support for relationships between graduate teacher credentials and effectiveness in the literature (see Chapter 2). The use of average years of teaching experience as a control helps address some of these concerns.

I present the results in Table 19 for math and ELA for both all students and the historically disadvantaged subgroup. These estimates are doubly robust using the following model, as a reminder from Chapter 3 (Methods):

$$y_i = a + \beta_1 + \beta_i X_i' + \varepsilon | w(x)$$

The vector of coefficients included in the model after weighting are enrollment, percentage of low-income enrollment, percentage of English learners, percentage of students on IEPs, student-teacher ratio, percentage of revenue that is local, percentage of Republican voters, adjusted salary, and average years of teaching experience in the district. Many of the significant relationships that appeared when using the previous techniques (OLS and spatial autoregression) are no longer present. As expected, the size of the estimates are small, considering that district-level controls that can explain much of the variation in student achievement.

There, however, a few significant findings in Table 19. One is the ability of prescribed negative consequences (present in 68 districts) to predict ELA achievement for historically disadvantaged students by 2.19%. In these districts with negative consequences, teachers who do not improve on student growth and proficiency measures do not receive an annual pay raise and eventually are subject to termination. While all districts have the ability to enact these policies under the Pennsylvania School Code as written in 2016-2017, only 68 chose to specify these procedures in the contract.

There are three other significant positive findings for the teacher qualification indicators: a 1.76% increase for all students in math proficiency on the PSSA in the 160 districts with teacher qualification items for teachers (such as mandatory prep time). There is also one significant and positive estimate on math (1.76%) for the seniority-based layoff treatment. I speculate that these 215 districts' students may benefit from seniority protections in harder-to-staff subjects such as mathematics. Teaching

effectiveness increases steadily over time (Papay & Kraft, 2016), and seniority protections may encourage teachers to improve (and give them more time to learn how to teach to the state test).

Table 19. Estimates of Contract Provisions on Achievement, Grades 3-8

	Math Effect (All)	ELA Effect (All)	Math Effect (Dis)	ELA Effect (Dis)
Strength (>0.50)	1.199	1.018	1.235	1.187
	(1.063)	(0.986)	(1.031)	(1.128)
Strength (<0.25)	-0.294	-0.341	-0.577	-0.441
	(0.858)	(0.650)	(0.815)	(0.705)
Classroom (>0)	0.032	0.473	0.586	0.458
	(1.296)	(1.224)	(1.196)	(1.311)
Teacher Qual. (>0.50)	1.756*	0.681	1.448	0.629
	(0.891)	(0.689)	(0.787)	(0.715)
National Board Bonus	1.703	0.820	1.750	0.820
	(0.976)	(0.710)	(0.975)	(0.873)
Max Class Size	1.081	1.357	1.470	1.200
	(1.368)	(1.297)	(1.261)	(1.429)
Avg Adj Salary ($76k)	1.101	0.197	1.016	0.536
	(0.799)	(0.597)	(0.747)	(0.731)
Negative Consequences	-0.392	0.662	0.760	2.190**
	(0.934)	(0.771)	(0.908)	(0.827)
Seniority Layoff	1.758*	0.836	1.311	1.089
	(0.836)	(0.667)	(0.757)	(0.695)
Mandatory Prep Time	0.460	0.436	0.587	0.731
	(0.815)	(0.670)	(0.750)	(0.736)
Seniority Hiring	0.155	-0.564	0.414	-0.819
	(1.226)	(0.947)	(1.070)	(0.986)
Longevity Bonus	1.222	-0.026	0.622	-0.818
	(2.092)	(1.374)	(2.055)	(1.540)
Republican Board	0.030	-0.338	-0.688	-1.139
	(0.894)	(0.637)	(0.824)	(0.733)
Majority Rep. District	0.223	-0.395	-0.525	-0.187
	(0.951)	(0.755)	(0.847)	(0.829)

Standard errors in parentheses. “All” refers to average student achievement in a district, “Dis” refers to average student achievement for disadvantaged subgroups only. There are 499 observations in each sample.
* $p < 0.05$, ** $p < 0.01$

Table 20 presents the results for the three high school Keystone examinations in Algebra, Biology, and Literature. There are a few significant results and larger standard errors. Two statistically significant results appear for historically disadvantaged students taking the Biology exam, suggesting that two contract provisions – higher salaries (+2.83%) and seniority-based hiring (+4.05%) – may be attracting or retaining more effective biology teachers, a notoriously hard-to-staff area (Ingersoll & Perda, 2010; Keefe, 2018). A third statistically significant result appears for Algebra test scores in districts that have a majority Republican voter registration (50% or greater, 203/499 Pennsylvania districts). 2.4% more students pass the Keystone Algebra exam in Republican districts than in Democratic ones, which was the only significant political predictor across any of the models.

Table 20. Estimates of Contract Provisions on Achievement, High School Tested Subjects

	Algebra (All)	Algebra (Dis)	Biology (All)	Biology (Dis)	Lit. (All)	Lit. (Dis)
Strength (>0.50)	-1.402	-1.537	-0.221	-0.708	0.630	0.861
	(1.532)	(1.932)	(1.514)	(1.929)	(1.210)	(1.612)
Strength (<0.25)	0.193	0.751	0.193	0.518	0.865	1.098
	(1.027)	(1.257)	(1.027)	(1.331)	(0.857)	(1.255)
Classroom (>0)	1.753	1.452	0.697	-0.626	2.768	3.541
	(1.777)	(2.253)	(1.716)	(2.307)	(1.431)	(1.933)
Teacher Qual. (>0.50)	0.217	0.015	0.664	0.690	-0.226	-0.974
	(0.991)	(1.263)	(1.031)	(1.286)	(0.863)	(1.209)
National Board Bonus	0.752	-1.251	0.552	-1.595	-0.415	-1.758
	(0.950)	(1.402)	(1.068)	(1.464)	(0.878)	(1.375)
Max Class Size	0.092	-0.013	-0.683	-2.535	0.253	0.750
	(1.681)	(2.385)	(1.928)	(2.676)	(1.556)	(2.214)
Top Avg Adj Salary ($76k)	0.576	1.697	1.109	2.834*	0.345	1.143
	(0.954)	(1.362)	(1.003)	(1.338)	(0.741)	(1.268)
Negative Consequences	-0.578	0.434	0.309	2.609	0.664	2.038
	(1.136)	(1.498)	(1.090)	(1.549)	(0.786)	(1.310)
Seniority Layoff	-0.081	0.047	-0.562	-0.093	-0.240	0.173
	(0.907)	(1.167)	(0.930)	(1.125)	(0.787)	(1.107)
Mandatory Prep Time	0.002	-0.092	0.588	0.536	-0.830	-0.886
	(1.021)	(1.366)	(1.093)	(1.635)	(0.902)	(1.495)
Seniority Hiring	1.182	1.194	2.396	4.048*	0.551	-0.679
	(1.265)	(1.731)	(1.241)	(1.751)	(0.953)	(1.639)
Longevity Bonus	3.495	1.018	3.326	-0.817	2.349	1.212
	(1.844)	(2.907)	(2.166)	(3.297)	(1.483)	(2.367)
Republican Board	0.868	-1.018	0.360	-1.379	1.202	0.233

	(0.904)	(1.231)	(0.890)	(1.255)	(0.724)	(1.201)
Majority Rep. District	2.404*	1.860	1.144	0.347	0.899	-0.340
	(0.976)	(1.330)	(0.978)	(1.374)	(0.738)	(1.226)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Importantly, none of the significant relationships are negative, which directly addresses a concern that teachers' rights and benefits come at the expense of student achievement. I find no evidence to support this hypothesis when using proficiency as an outcome. However, I do find evidence to support providing math teachers in grades 3-8 with cumulative contractual benefits (as a reminder, this construct includes longevity bonuses, NBC bonus, professional growth plans, preparation time, and all of the seniority-based hiring and firing provisions). These items in isolation do not demonstrate relationships to student achievement, but they do when combined into a construct with high internal validity. Here again this result may speak to the difficulty of staffing math teacher positions compared to ELA. I also find evidence to support the efficacy of negative consequences for ELA teachers of disadvantaged students, as well as higher salaries for high school biology teachers. I caution against overinterpreting these findings because of the limitations of the study design and the small magnitude of these estimates. Still, it is remarkable to find even these small estimates considering the robustness of the models.

There are many limitations when using these techniques, but I argue that they are the best available considering the single year of data and the relatively small sample size (n = 499). Furthermore, my examination of the 105 new contracts in 2016-2017 to their older versions shows great stability in their content over time. The results of PSW cannot account for endogeneity and unobserved variables, but they can provide more rigorous estimates than OLS and GS2SLS. Table 21 demonstrates how these estimates shift

depending upon the technique. The positive estimate for historically disadvantaged biology test takers is the least sensitive to changes in the approach.

Table 21: Comparison of Significant Estimates

	OLS	GS2SLS	GBM
Teacher Qualification Items, Math (All)	3.205	2.783	1.756*
	(1.726)	(1.716)	(0.827)
Negative Consequences, ELA (Disadvantaged)	-0.145	1.361	2.190**
	(0.738)	(0.834)	(0.891)
Republican Voting Majority, Algebra (All)	0.289	0.572	2.404*
	(1.179)	(0.903)	(0.976)
Seniority Hiring, Biology (Disadvantaged)	3.230*	3.144*	4.048*
	(1.345)	(1.461)	(1.751)

In Table 22, I present the results of the graduation rate models. In contrast to Table 20, all of the significant results are negative rather than positive. Graduation rates are an increasingly troubled measure of student success considering the skewness of the data (with a mean of 91% unweighted, or 88% weighted by enrollment) and the problematic use of credit recovery in large urban school districts (Malkus & Cummings, 2018). Only thirteen districts out of 499 had a graduation rate below 70% in 2015-2016. Still, I treat graduation rates as another measure of student success despite the admittedly low bar. I examine only the White subgroup because many districts did not have sufficient numbers to qualify for other types of racial, ethnic, and economic subgroup reporting.

In Table 22, the teacher qualification construct that positively predicts math achievement in grades 3-8 becomes negative and much larger in magnitude. I can only speculate the reasons behind this trend and highlight it as an important area for qualitative inquiry. It may be the case that the cumulative seniority protections included in the

teacher qualification indicators cause teachers to disinvest from pushing struggling students across the graduation finish line. Goodson, Moore, and Hargreaves (2006) claim that teachers retain "historically situated missions formed decades ago that teachers have carried with them throughout their careers," and this mission might include a sense of academic gatekeeping (p. 46). Veteran teachers may be more resistant to the growing pressure to raise graduation rates irrespective of students' content knowledge (Day & Gu, 2009). In these doubly robust models, the teacher experience average is also negative, though it is not statistically significant.

Table 22. Estimates of Contract Provisions on Graduation Rates

	Graduation Rate (All)	Graduation Rate (Male)	Graduation Rate (Female)	Graduation Rate (White)
Strength (>0.50)	-0.650	-0.550	-0.742	-0.626
	(1.031)	(1.267)	(0.925)	(1.082)
Strength (<0.25)	0.867	1.156	0.563	0.671
	(0.507)	(0.634)	(0.496)	(0.514)
Classroom (>0)	-1.614	-1.567	-2.213*	-2.275
	(1.156)	(1.555)	(0.989)	(1.305)
Teacher Qual. (>0.50)	-5.681**	-7.831**	-3.058*	-5.499**
	(1.389)	(1.929)	(1.454)	(1.571)
National Board Bonus	0.800	0.900	0.751	0.452
	(0.765)	(0.921)	(0.679)	(0.871)
Max Class Size	-1.533	-0.877	-2.167	-2.451
	(1.050)	(1.253)	(1.128)	(1.336)
Avg Adj Salary ($76k)	-0.402	-0.473	-0.092	-0.130
	(0.649)	(0.777)	(0.661)	(0.707)
Negative Consequences	0.168	0.098	0.169	-0.005
	(0.620)	(0.788)	(0.540)	(0.554)
Seniority Layoff	0.251	-0.078	0.564	-0.147
	(0.558)	(0.683)	(0.535)	(0.574)
Mandatory Prep Time	-1.153*	-1.497*	-0.755	-0.750
	(0.490)	(0.607)	(0.493)	(0.528)
Seniority Hiring	0.117	-0.017	0.275	0.559
	(0.756)	(0.847)	(0.751)	(0.741)
Longevity Bonus	-0.306	-0.227	-0.338	-0.844
	(0.661)	(0.815)	(0.639)	(0.692)
Republican Board	0.878	0.313	1.497	0.984
	(0.486)	(0.615)	(0.455)	(0.531)
Majority Rep. District	-0.198	-0.629	0.180	-0.317
	(0.464)	(0.572)	(0.497)	(0.436)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

CHAPTER 6: CONCLUSIONS

To return to the question asked in the title: are collective bargaining reforms worth the fight? That is, should reformers "take on" unions in order to raise student achievement? My study suggests that, at least in Pennsylvania, the answer is "no" for students in grades 3-8. In the elementary and middle grades, teachers and students seem to benefit from an array of teacher qualification provisions, though overall results are null or quite small in magnitude. But perhaps the answer is "yes" at the high school level, if we are considering graduation rates and the need to incentivize the hiring of more highly effective biology teachers. Drawing upon my own professional experience as a 12th grade ELA teacher, where I was directed to pass students at any cost, I remain skeptical of studies that rely on graduation rates alone. I find the biology results much more persuasive and consistent across the various models. During the time frame of this study, the biology tests also were not a high school graduation requirement, so there was less pressure for teachers to teach to the test.

My findings are somewhat in keeping with the four previously studied states, as the majority of my results are null with some important differences. The largest positive relationships to student achievement are between teacher salaries (which may be endogenous, as teachers flock to higher paying districts that already have higher student achievement) and lower student-teacher ratios. Salaries are most likely a function of the ability of districts to leverage local property taxes amid fiscal constraints. State-level funding structures favor rural districts over urban ones (Steinberg & Quinn, 2015), and there is an enormous range across Pennsylvania in terms of that percentage of local

property tax that accounts for school funding – from only 10% of overall district expenditures in Duquense City to 86% in wealthy Lower Merion. The remainder of these district budgets are covered by state and federal funds. Considering these realities, the most effective teachers may be sorting themselves into districts with the highest pay, and this environment may also increase these teachers' effectiveness. This effectiveness may not be portable and may also depend upon the context of these wealthier districts – or smaller districts with low student-teacher ratios (Papay & Kraft, 2016).

While effectiveness may not be portable, mobility for veteran teachers may be relatively frictionless. The spatial dependence of contracts across the state illustrates that teachers can change their employer without a loss of contractual benefits. Districts are matching each other not only in terms of salary but in also in terms of lifetime benefits that take the form of seniority protections. Ironically, hard-to-staff districts such as Philadelphia have removed these protections in the name of reform. The lack of policy diffusion from urban areas suggests that suburban districts will continue to pay a premium for highly educated teachers, as wealthier parents may implicitly value these credentials – particularly when proficiency on a state test is all but a given. Proficiency itself is dependent upon neighboring performance. Districts seem to care more about their neighboring competitors than the latest education reform.

However, paying a premium salary in a hard-to-staff subject such high school biology may be a worthwhile option for policymakers to consider. Other alternatives – such as loan forgiveness programs – have also been effective at improving retention and achievement (Feng & Sass, 2018). Under current collective bargaining rules, such a

targeted proposal would no doubt be politically challenging. Differential teacher pay for different subject areas conflicts with bedrock principles of collective bargaining, but it could be a significant leverage point for improving the status of the teaching profession. It makes little sense to raise teacher pay for subjects for which there are hundreds of applicants, such as high school English (Ingersoll & Merrill, 2017), and this study should further the dialogue concerning differentiated pay for different grades and subjects. One potential loophole is that offering graduate credit bonuses could reward those who have earned advanced degrees in engineering and computer science without penalizing teachers with master's degrees in the humanities.

On an additional positive note for proponents of seniority protections, the much-maligned "last-in, first-out" rule is not negatively related to elementary and middle grade proficiency. In the case of math proficiency in grades 3-8, the relationship between achievement and this seniority protection was, in fact, significant and positive. A major caveat to this finding is that while many contracts did not specifically enumerate this protection, it may still be in effect by default in those districts as a result of the Pennsylvania School Code. Still, enumerating this protection indicates the strength of the local union and the values of the district. And it takes years for contract reforms to improve teacher quality at the cost of institutional knowledge and teacher turnover (Morel, 2018; Schueler, Goodman, & Deming, 2017).

The negative finding for high school graduation rates, however, is an area of concern. I offer two possible explanations based on anecdotal teaching experience. Veteran teachers may be holding students to a higher standard, whereas novice teachers

may be more susceptible to administrative pressure to graduate seniors and improve the district's graduation metric. This phenomenon, while not well-researched, was my experience in Lawrence, and it is a growing national concern that is difficult to document and prove (Dynarski, 2018; Malkus & Cummings, 2018). Pushing seniors out the door would certainly be another example of the kids-on-the-bubble behaviors that have plagued accountability systems based on a simple cut-off score (e.g., Booher-Jennings, 2005). Graduation is a stark, binary outcome, and it is far easier to manipulate than a standardized test. Qualitative work is needed in this area, as schools continue to be lauded for improving their graduation rates without much critical inquiry, though scandals have emerged in districts with strong accountability pressures such as Washington, D.C (Stein, 2018).

While I cannot make causal claims in this study, the lack of significant relationships among most of the contract provisions and student achievement suggests that weakening union protections may do little to improve the quality of the teaching workforce or raise student achievement. The rise of teacher strikes and other activist movements suggests that we have reached an inflection point, as political attention turns to more material concerns that students face – poverty, hunger, homelessness, and inadequate health care. No matter how complex the model in this study, family income remains the strongest determinant, and the political hypotheses that drove this work have largen been disproven. Conservatives may have contractual preferences, but these do not result in improved student outcomes. When looking at contracts and student outcomes,

districts are more attentive to what their neighbor is offering than to political conditions within the community.

Weakening union protections may also adversely impact teacher pay, as teacher pay is far lower in states without collective bargaining (Paglayan, 2018), which in turn may worsen student achievement. When placed in the context of the literature on teacher evaluation system modifications (Donaldson & Woulfin, 2018; Firestone, 2014; Stecher et al., 2018), such highly contentious changes may be a waste of political capital. In the studies referenced throughout, achievement effects are either nonsignificant or small and often come at great political cost.

Education reformers seeking to overturn union regulations appear to be operating on a set of assumptions that have an increasingly weak empirical basis, though I highlight studies such as Lovenheim & Willén (2019) that have plausibly causal claims that unions can be harmful. Ultimately, I find no significant negative relationships between any items and student proficiency and a few positive relationships. Efforts in the political sphere have already moved on to more fundamental concerns of student and staff health and safety.

The significant relationships to contract items or constructs that do appear occur in teacher qualification domains. The teacher qualification construct includes bonuses for teachers that occur as part of a package of credentialing incentives, as opposed to a focus on class size restrictions or teacher evaluations. Ironically, the classroom construct covers two areas of far greater legislation, research, and interest over the past twenty years. In one instance, however, negative consequences for poor ratings did predict significantly

higher ELA achievement for historically disadvantaged students, though only when using the propensity score weighting approach. The teaching of English may be more amenable to accountability pressures and teacher expertise, as ELA curricula can be overly vague in the absence of an aligned state test or standards-based curricula (Conley, 2011; Edgerton, 2019a; Strauss, 2016).

My hope is that these findings, when placed within the larger literature, dampen the notion that a comprehensive weakening of collective bargaining provisions may benefit students. Policy solutions demand a scalpel – such as higher pay for math and science teachers – rather than an anti-union sledgehammer. The graduation rate, which is extremely high in all but a few districts, should not carry the same weight as achievement in policy discussions, as the validity of this measure is highly suspect. And achievement itself is spatially dependent, suggesting that there is a sharing of resources and perhaps even staff, over time, across district lines. In Pennsylvania, some districts even cross multiple county lines. Testing a geospatial theoretical framework further emphasizes how proficiency reflects a built-in advantage that is geographic as well as income-based.

An important caveat is that this study ignores some of the most expensive contractual guarantees, and it is not a cost-benefit analysis. Teacher strikes continue to occur across the country but with very different demands. In Denver, the union contested noncompetitive pay and a glut of administrators that they perceived to be superfluous (Campbell, 2019). Chicago teachers, in contrast, lobbied for social workers and nurses in their schools (Smith & Davey, 2019). This dissertation does not compare a set of fiscal priorities, which is the reality of day-to-day school district governance (Dahill-Brown,

2019). The bulk of what makes up a contract – health insurance, retirement, and other benefits – are the more likely domains for future political battles as districts struggle to pay for ballooning pension costs (Glaeser & Ponzetto, 2014). The coronavirus pandemic will no doubt make these concerns more acute, with education reform fading into the background of more pressing material needs.

But the fact remains that there is a specific type of teacher shortage in Pennsylvania – particularly in urban areas, in STEM, for English learners, for students with disabilities, and for historically disadvantaged students. There is no teacher shortage in the humanities. Teachers, particularly in STEM subjects, remain dramatically undercompensated relative to their educated peers in higher-paying professions. This disparity can only be partially explained by the reality that more of teacher compensation comes in the form of benefits, rather than wages, when compared to their private sector counterparts (Keefe, 2018). Additionally, these generous benefits are steadily decreasing across the United States and across sectors, and they will most likely continue to do so. Benefit reductions may be negatively impacting the quality of teachers entering the workforce; the number of teachers graduating from teacher training programs in Pennsylvania plummeted by 63% from 2013 to 2015 (Keefe, 2018).

Considering my findings and the existing literature, I conclude that the most problematic aspects of collective bargaining in terms of student achievement are not the minutiae such as seniority preference within the provisions themselves, which consume much political energy. The core problem, as I see it, is the lack of salary differentiation across subject areas. Higher salaries for ELA teachers are not significantly related to

student achievement in this study when controlling for spatial dependence. This finding is important as not considering spatial controls may cause researchers to overstate the ability of teacher salaries to increase teacher quality. I also emphasize that the narrow range of outcomes in this study – student achievement and graduation rates in a single year – are only a small step towards more plausibly causal claims, and towards looking at other measures of student health and well-being. New statewide longitudinal data systems are now collecting information on more than just test scores.

Still, higher average math teacher salaries in a district demonstrate significant relationships to achievement. I provide suggestive evidence supporting differential pay across difficult-to-staff subject areas, considering that Pennsylvania has a math teacher shortage and poorer districts may be more likely to hire and retain lower quality math teachers when compared to ELA (Keefe, 2018). Prior studies have also found higher salaries to be helpful at retaining teachers, particularly if teachers in Pennsylvania are willing to move beyond their neighboring districts (Feng, 2014). Throughout this study, the spatial dependency of contracts, wages, and key items suggests that districts do worry that teachers in higher demand will move.

As for ELA teacher policy, it may make more sense to focus on retention in order boost the average years of experience for a school district, considering the significant relationship between experience and student achievement. English teaching may take more time to master considering the wide range of literature, tasks, curricula, and assessments that may be administered, whereas STEM subjects may have a more clearly defined scope and sequence (Achieve, 2018; Edgerton, 2019a; Strauss, 2016). However,

ELA teachers may be more likely to stay in the classroom because of the lack of higher-paying private sector jobs compared to their STEM colleagues. Few reformers make the argument that humanities teachers without additional licensures (e.g., for English learners and students with disabilities) should be paid more considering their oversupply, and certainly this study does not support that line of thinking.

Finally, I turn to the question of partisan school governance and find a few significant relationships, all of which occurred in STEM subjects (either in math grades 3-8, in high school algebra, or in biology) depending upon the model and the technique. These findings provide suggestive evidence that Republicans – both voters and school boards – may emphasize these subjects over the humanities, which matches the political rhetoric of the moment. Conservatives are more likely to support a vocational view of schooling with clear economic returns on public investment (Tetlock et al., 2013). But politics does not appear to matter as much as proximity; most of the political findings become nonsignificant when incorporating spatial weights.

I should pause here and note that I taught high school ELA for several years and do not intend to disparage the incredible effort of ELA educators. But I find in the data a compelling story, where math teachers (who are in higher demand) may be more responsive to financial incentives, and ELA teachers may be more responsive to negative accountability. ELA teachers may change their instruction to match the state test (see Edgerton & Desimone, 2018, for an example of how standards-based policy can positively relate to ELA instruction in another state, Texas).

These findings, I caution, are still quite small in magnitude. The largest finding for graduation rates is a highly skewed outcome that has little bearing on policy for affluent districts, where an overwhelming majority of students graduate. Considering these limitations, it is important to place this work in its historical context. CBAs are remarkably resistant to change considering their decades-long histories, stretching back to the very first days of collective bargaining in the 1970's (Ingle & Wisman, 2018). During my reading of the contracts, I uncovered provisions dated from the 1970's that are still in effect, and I found only a handful of changes for new contracts that took effect in 2016-2017. This stability is consistent with the limited literature available (Ingle & Wisman, 2018).

Looking across my findings, I suggest that the conservative emphasis on weakening collective bargaining (as in Wisconsin under Governor Scott Walker, to cite the most prominent example) is an inefficient use of political capital, if the goal is to raise student achievement. Districts are more likely to follow their neighbor's lead instead of the state capital's agenda. However, as pension costs balloon, amending CBAs may become necessary to keep school districts solvent. But these battles should not be waged over the issues that have captured the most attention from education reformers when placing this study within the broader literature. Aside from high school graduation rates, I find little empirical basis for the disproportionate political focus on the removal of seniority rights. Pennsylvania's children would be better served instead by focusing on grossly inequitable funding structures, which may incentivize the best teachers to leave those districts that need them most.

APPENDIX

Table 23. OLS Results for Math and ELA Achievement Weighted by Enrollment

	Math % Proficient	Math % Proficient	Math % Proficient	ELA % Proficient	ELA % Proficient	ELA % Proficient
Strength	0.485			1.294		
	(1.950)			(1.379)		
Classroom		-2.557			-1.797	
		(1.437)			(1.341)	
Teacher Qual.			1.067			-0.223
			(1.846)			(1.377)
PPE ($1000s)	-0.272*	-0.181	-0.292*	-0.414**	-0.347**	-0.406**
	(0.123)	(0.119)	(0.126)	(0.108)	(0.107)	(0.108)
% Low-Inc.	-0.538**	-0.534**	-0.539**	-0.516**	-0.513**	-0.516**
	(0.034)	(0.034)	(0.034)	(0.027)	(0.028)	(0.027)
% IEP	-0.597**	-0.581**	-0.594**	-0.568**	-0.556**	-0.567**
	(0.148)	(0.147)	(0.148)	(0.099)	(0.100)	(0.099)
% ELL	-0.330	-0.366	-0.326	-0.457**	-0.483**	-0.458**
	(0.272)	(0.266)	(0.276)	(0.137)	(0.132)	(0.137)
S-T Ratio	-1.205**	-1.005**	-1.241**	-1.376**	-1.217**	-1.342**
	(0.278)	(0.287)	(0.298)	(0.253)	(0.248)	(0.252)
% Local Tax	0.101*	0.102*	0.103*	0.095**	0.095**	0.094**
	(0.042)	(0.041)	(0.041)	(0.027)	(0.027)	(0.027)

Rep. Board	-0.031	-0.096	0.012	0.341	0.270	0.294
	(1.136)	(1.113)	(1.145)	(0.900)	(0.894)	(0.901)
Suburban	-1.573	-2.327	-1.505	0.352	-0.221	0.274
	(2.461)	(2.524)	(2.497)	(1.589)	(1.589)	(1.599)
Town	0.277	-0.555	0.394	1.815	1.099	1.598
	(2.561)	(2.565)	(2.600)	(1.926)	(1.916)	(1.929)
Rural	-3.235	-4.022	-3.140	-0.765	-1.437	-0.959
	(2.495)	(2.501)	(2.535)	(1.792)	(1.788)	(1.811)
Sal. ($1000s)	0.150**	0.147**	0.150**	0.138**	0.137**	0.139**
	(0.051)	(0.050)	(0.051)	(0.037)	(0.036)	(0.036)
Years of Exp.	0.110	0.090	0.116	0.244	0.227	0.238
	(0.261)	(0.261)	(0.262)	(0.190)	(0.191)	(0.193)
Constant	85.573**	82.309**	85.804**	105.045**	102.871**	105.172**
	(7.100)	(7.306)	(7.116)	(6.535)	(6.709)	(6.527)
N	499	499	499	499	499	499
R^2	0.849	0.850	0.849	0.904	0.904	0.904
pseudo R^2						

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Table 24. OLS Results for Historically Disadvantaged Math and ELA Achievement Weighted by Enrollment

	Math % Proficient	Math % Proficient	Math % Proficient	Math % Proficient	Math % Proficient	Math % Proficient
Strength	1.195			2.210		
	(1.827)			(1.619)		
Classroom		1.208			1.676	
		(1.699)			(1.866)	
Teacher Qual.			1.975			1.733
			(1.792)			(1.567)
PPE ($1000s)	-0.088	-0.089	-0.099	-0.126	-0.131	-0.144
	(0.130)	(0.131)	(0.127)	(0.118)	(0.116)	(0.115)
% Low-Inc.	-0.170**	-0.170**	-0.171**	-0.183**	-0.183**	-0.184**
	(0.032)	(0.032)	(0.032)	(0.029)	(0.029)	(0.029)
% IEP	-0.462**	-0.469**	-0.460**	-0.579**	-0.588**	-0.576**
	(0.129)	(0.130)	(0.130)	(0.106)	(0.108)	(0.107)
% ELL	-0.481*	-0.476*	-0.479*	-0.577**	-0.568**	-0.571**
	(0.216)	(0.217)	(0.219)	(0.111)	(0.113)	(0.114)
S-T Ratio	-0.726*	-0.713*	-0.722*	-0.822**	-0.805**	-0.821**
	(0.289)	(0.292)	(0.290)	(0.273)	(0.271)	(0.273)
% Local Tax	0.059	0.059	0.063	0.065*	0.065*	0.069*
	(0.038)	(0.038)	(0.038)	(0.029)	(0.029)	(0.029)
Rep. Board	-0.210	-0.154	-0.082	0.129	0.184	0.192
	(1.078)	(1.077)	(1.101)	(0.961)	(0.994)	(0.975)
Suburban	-1.401	-1.405	-1.442	-0.237	-0.219	-0.225
	(2.664)	(2.671)	(2.668)	(1.870)	(1.889)	(1.902)

Town	0.411	0.353	0.465	1.283	1.169	1.260
	(2.577)	(2.570)	(2.592)	(2.040)	(2.053)	(2.079)
Rural	-2.026	-2.089	-2.009	-1.107	-1.219	-1.146
	(2.547)	(2.546)	(2.558)	(1.982)	(2.003)	(2.024)
Sal. ($1000s)	0.148**	0.145**	0.143**	0.152**	0.149**	0.150**
	(0.046)	(0.046)	(0.047)	(0.040)	(0.040)	(0.041)
Years of Exp.	0.095	0.105	0.113	0.220	0.232	0.231
	(0.244)	(0.242)	(0.247)	(0.218)	(0.221)	(0.221)
Constant	44.836**	45.267**	44.281**	63.925**	64.777**	63.972**
	(7.083)	(6.878)	(7.092)	(7.237)	(7.115)	(7.335)
N	499	499	499	499	499	499
R^2	0.849	0.850	0.849	0.904	0.904	0.904
pseudo R^2						

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Table 25. OLS Results for High School Student Achievement Weighted by Enrollment, All and Historically Disadvantaged (Dis.)

	Algebra (All)	Algebra (Dis.)	Biology (All)	Bio (Dis.)	Literature (All)	Literature (Dis.)
Strength	0.015	-2.400	1.633	-0.643	0.466	-1.438
	(1.950)	(2.650)	(2.040)	(2.788)	(1.734)	(2.509)
PPE ($1000s)	-0.078	0.140	-0.226	-0.051	-0.194	-0.054
	(0.170)	(0.227)	(0.165)	(0.229)	(0.128)	(0.182)
% Low-Inc.	-0.450**	-0.227**	-0.506**	-0.303**	-0.347**	-0.202**
	(0.052)	(0.060)	(0.050)	(0.060)	(0.046)	(0.058)
% IEP	-0.456**	-0.437*	-0.603**	-0.537**	-0.438**	-0.473*
	(0.175)	(0.212)	(0.160)	(0.207)	(0.149)	(0.200)
% ELL	-0.437	-0.255	-0.513	-0.387	-0.452	-0.229
	(0.413)	(0.442)	(0.348)	(0.391)	(0.331)	(0.394)
S-T Ratio	-0.795	-0.744	-1.214**	-1.314*	-0.830*	-1.104*
	(0.422)	(0.518)	(0.412)	(0.520)	(0.341)	(0.465)
% Local Tax	0.043	-0.028	0.049	-0.048	0.054	-0.017
	(0.058)	(0.067)	(0.053)	(0.066)	(0.046)	(0.060)
% Republican	0.126*	0.064	0.093*	0.037	0.067	-0.018
	(0.051)	(0.059)	(0.045)	(0.055)	(0.038)	(0.051)
Suburban	-1.364	1.279	-2.188	0.304	-0.294	2.398
	(3.630)	(3.853)	(2.695)	(3.198)	(2.489)	(3.215)
Town	1.444	3.139	2.012	3.764	1.777	4.047
	(3.815)	(4.129)	(2.786)	(3.186)	(2.673)	(3.515)

Rural	-1.933	0.363	-1.894	0.568	-0.354	2.069
	(3.706)	(3.988)	(2.707)	(3.182)	(2.553)	(3.397)
Sal. ($1000s)	0.109	0.149*	0.084	0.143	0.034	0.064
	(0.063)	(0.075)	(0.060)	(0.076)	(0.053)	(0.071)
Years of Exp.	0.066	-0.348	0.519	0.004	0.296	-0.092
	(0.312)	(0.407)	(0.282)	(0.375)	(0.251)	(0.360)
Constant	97.668**	73.409**	105.360**	85.851**	105.918**	95.270**
	(9.009)	(11.509)	(8.845)	(12.345)	(6.530)	(10.161)
N	499	499	499	499	499	499
R^2	0.849	0.850	0.849	0.904	0.904	0.904
pseudo R^2						

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Table 26. Balance of the Treatment and Comparison Groups, Strength > 0.25

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	18.81	18.28	0.5570	17.34	0.08
Enrollment (1000s)	11.42	4.36	0.2460	2.80	0.16
Low-Inc Percent	45.77	43.92	0.7180	43.09	0.57
IEP Percent	19.71	19.05	0.6230	18.12	0.20
ELL Percent	2.61	2.61	1.0000	1.31	0.05
S.T. Ratio	14.36	14.35	0.9880	14.00	0.47
Local Rev Percent	59.00	58.01	0.8200	51.34	0.06
Republican Percent	32.38	37.19	0.1050	46.22	0.00
Adj. Salary (1000s)	75.09	73.58	0.4580	70.07	0.01
Average Exp. (Years)	14.01	14.01	0.9880	14.27	0.54

Table 27. Balance of the Treatment and Comparison Groups, Classroom

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	18.47	18.46	0.9930	17.32	0.10
	7.83	3.95	0.2990	2.81	0.18
Low-Inc Percent	46.59	43.18	0.4230	42.95	0.34
IEP Percent	19.37	18.74	0.4890	18.10	0.13
ELL Percent	1.91	2.06	0.7820	1.32	0.18
S.T. Ratio	14.05	14.04	0.9720	14.01	0.90
Local Rev Percent	58.34	57.75	0.8600	51.16	0.02
Republican Percent	32.90	37.66	0.0460	46.60	0.00
Adj. Salary (1000s)	72.94	72.33	0.7380	70.08	0.08
Average Exp. (Years)	14.05	14.12	0.8500	14.28	0.47

Table 28. Balance of the Treatment and Comparison Groups, Teacher Qualification > 0.50

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	17.87	17.84	0.9540	17.18	0.04
Enrollment (1000s)	4.29	2.99	0.1370	2.63	0.05
Low-Inc Percent	45.00	44.24	0.7400	42.35	0.14
IEP Percent	18.78	18.34	0.3040	17.90	0.02
ELL Percent	1.73	1.61	0.7490	1.19	0.06
S.T. Ratio	14.03	14.04	0.9520	14.00	0.86
Local Rev Percent	50.71	52.17	0.4960	52.12	0.45
Republican Percent	38.87	41.29	0.0820	48.86	0.00
Adj. Salary (1000s)	71.00	70.60	0.6620	69.94	0.21
Average Exp. (Years)	14.00	14.17	0.3790	14.38	0.03

Table 29. Balance of the Treatment and Comparison Groups, National Board Cert

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	18.21	17.95	0.6060	17.28	0.05
Enrollment (1000s)	6.38	3.88	0.2230	2.68	0.07
Low-Inc Percent	37.29	38.98	0.5580	44.09	0.01
IEP Percent	17.18	17.66	0.3380	18.34	0.01
ELL Percent	2.13	2.04	0.8280	1.25	0.02
S.T. Ratio	14.12	14.17	0.8180	14.00	0.56
Local Rev Percent	62.34	59.52	0.2500	50.06	0.00
Republican Percent	41.45	42.87	0.4180	46.26	0.00
Adj. Salary (1000s)	71.97	71.84	0.9180	70.03	0.09
Average Exp. (Years)	14.13	14.11	0.9160	14.28	0.50

Table 30. Balance of the Treatment and Comparison Groups, Max Class Size

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	18.58	18.15	0.6250	17.34	0.12
Enrollment (1000s)	9.75	4.05	0.2710	2.82	0.18
Low-Inc Percent	48.09	46.20	0.7100	42.95	0.27
IEP Percent	19.39	19.13	0.8190	18.12	0.23
ELL Percent	2.47	2.46	0.9850	1.31	0.04
S.T. Ratio	14.23	14.28	0.9230	14.00	0.60
Local Rev Percent	59.63	56.97	0.4900	51.24	0.01
Republican Percent	31.21	35.70	0.0840	46.40	0.00
Adj. Salary (1000s)	73.26	72.27	0.6440	70.13	0.10
Average Exp. (Years)	13.72	13.82	0.7960	14.29	0.13

Table 31. Balance of the Treatment and Comparison Groups, Negative Consequences

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	18.21	18.07	0.7650	17.27	0.03
Enrollment (1000s)	6.08	3.86	0.2580	2.70	0.08
Low-Inc Percent	36.18	39.10	0.2360	44.31	0.00
IEP Percent	17.93	18.01	0.8470	18.23	0.48
ELL Percent	1.86	1.99	0.6870	1.29	0.05
S.T. Ratio	14.28	14.11	0.4470	13.97	0.15
Local Rev Percent	62.78	60.08	0.2550	49.91	0.00
Republican Percent	43.06	44.03	0.5580	46.04	0.06
Adj. Salary (1000s)	72.36	71.74	0.6450	69.95	0.05
Average Exp. (Years)	14.12	14.14	0.9120	14.28	0.42

Table 32. Balance of the Treatment and Comparison Groups, Seniority Layoff

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	18.03	17.57	0.1860	16.92	0.00
Enrollment (1000s)	3.69	3.08	0.3820	2.76	0.16
Low-Inc Percent	44.03	45.09	0.6110	42.58	0.36
IEP Percent	18.66	18.41	0.5040	17.83	0.02
ELL Percent	1.65	1.63	0.9720	1.15	0.05
S.T. Ratio	13.80	13.94	0.3830	14.17	0.01
Local Rev Percent	51.31	51.20	0.9540	51.93	0.72
Republican Percent	41.58	43.29	0.2000	48.70	0.00
Adj. Salary (1000s)	71.03	70.24	0.3420	69.72	0.09
Average Exp. (Years)	14.06	14.11	0.7800	14.41	0.03

Table 33. Balance of the Treatment and Comparison Groups, Mandatory Prep Time

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	17.34	17.39	0.8870	17.57	0.59
Enrollment (1000s)	3.48	2.67	0.0670	2.19	0.00
Low-Inc Percent	42.81	43.48	0.7420	44.40	0.32
IEP Percent	18.13	18.28	0.7580	18.35	0.56
ELL Percent	1.48	1.60	0.8350	1.02	0.07
S.T. Ratio	14.09	14.11	0.9310	13.77	0.06
Local Rev Percent	52.78	51.24	0.4870	48.26	0.02
Republican Percent	44.62	45.66	0.4790	48.75	0.00
Adj. Salary (1000s)	70.58	70.37	0.8060	69.38	0.18
Average Exp. (Years)	14.19	14.25	0.7870	14.49	0.09

Table 34. Balance of the Treatment and Comparison Groups, Seniority Hiring

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	17.40	17.11	0.4910	17.40	0.99
Enrollment (1000s)	4.48	3.23	0.4370	2.90	0.32
Low-Inc Percent	44.98	44.24	0.7200	42.85	0.27
IEP Percent	18.66	18.27	0.4050	18.09	0.18
ELL Percent	1.10	1.31	0.4620	1.42	0.23
S.T. Ratio	14.38	14.31	0.7000	13.94	0.02
Local Rev Percent	49.66	50.34	0.7600	52.06	0.26
Republican Percent	41.11	42.44	0.3270	46.54	0.00
Adj. Salary (1000s)	72.16	71.79	0.7750	69.91	0.05
Average Exp. (Years)	14.32	14.32	0.9870	14.25	0.70

Table 35. Balance of the Treatment and Comparison Groups, Longevity Bonus

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	17.40	17.45	0.8840	17.40	0.99
Enrollment (1000s)	2.86	2.86	0.9950	3.25	0.40
Low-Inc Percent	44.85	43.98	0.6830	42.75	0.29
IEP Percent	18.00	18.00	0.9960	18.24	0.57
ELL Percent	1.34	1.23	0.7020	1.37	0.92
S.T. Ratio	13.85	13.80	0.7780	14.06	0.22
Local Rev Percent	50.10	50.05	0.9840	52.10	0.34
Republican Percent	44.05	44.55	0.7520	46.08	0.17
Adj. Salary (1000s)	70.60	70.59	0.9920	70.19	0.68
Average Exp. (Years)	14.10	14.23	0.5690	14.30	0.31

Table 36. Balance of the Treatment and Comparison Groups, Republican Board

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	17.61	17.14	0.9460	17.78	0.05
Enrollment (1000s)	3.20	2.76	0.3600	3.10	0.85
Low-Inc Percent	40.36	41.04	0.6530	47.67	0.00
IEP Percent	19.71	19.05	0.6230	18.12	0.20
ELL Percent	1.21	1.14	0.7160	1.61	0.10
S.T. Ratio	14.01	14.15	0.4560	14.02	0.96
Local Rev Percent	52.27	52.21	0.9760	50.70	0.36
Adj Salary (1000s)	70.11	70.28	0.8430	70.55	0.60
Average Exp. (Years)	14.39	14.16	0.1660	14.05	0.03

Table 37. Balance of the Treatment and Comparison Groups, Majority Rep. District

	tx.mn	ct.mn	p	ct.mn	p
PPE (1000s)	17.02	17.04	0.9470	17.66	0.03
Enrollment (1000s)	2.38	2.60	0.2950	3.70	0.01
Low-Inc Percent	41.59	41.83	0.8510	44.31	0.06
IEP Percent	18.01	18.30	0.4040	18.31	0.37
ELL Percent	0.95	0.92	0.8540	1.65	0.00
S.T. Ratio	13.78	13.94	0.3670	14.17	0.01
Local Rev Percent	49.59	50.59	0.5760	53.09	0.04
Adj Salary (1000s)	68.95	69.75	0.3230	71.20	0.00
Average Exp. (Years)	14.49	14.40	0.5750	14.10	0.01

BIBLIOGRAPHY

Achieve. (2018). *Accountability in state ESSA plans*. Retrieved from https://states.achieve.org/essa-tracker

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 54-76.

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.

Anzia, S. F., & Moe, T. M. (2014). Collective bargaining, transfer rights, and disadvantaged schools. *Educational Evaluation and Policy Analysis*, *36*(1), 83-111.

Anzia, S. F., & Moe, T. M. (2016). Do politicians use policy to make politics? the case of public-sector labor laws. *American Political Science Review*, *110*(4), 763-777.

Associated Press (2019). Pay disparity hampers teacher retention in small districts. *The Spokesman Review*.

Au, W. (2016). Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism. *Educational Policy*, *30*(1), 39-62.

Babcock, L., Engberg, J., & Greenbaum, R. (2005). Wage spillovers in public sector contract negotiations: The importance of social comparisons. *Regional Science and Urban Economics*, *35*(4), 395-416.

Ballou, D. (2000a). Contractual constraints on school management: Principals' perspectives on teacher contracts. In D. Ravitch & J. Viteritti (Eds.), *City schools: Lessons from New York* (pp. 89-116). Baltimore, MD: The Johns Hopkins University Press.

Ballou, D. (2000b). *Teacher contracts in Massachusetts*. Boston, MA: Pioneer Institute for Public Policy Analysis.

Barnett, J. (2018 Jun). *Government union efforts to restrict workers' rights.* Harrisburg, PA: Commonwealth Foundation.

Bartanen, B., Grissom, J. A., Joshi, E., & Meredith, M. (2018). Mapping inequalities in local political representation: Evidence from Ohio school boards. *AERA Open*, *4*(4): 1-19.

Bau, N., & Das, J. (2017). *The misallocation of pay and productivity in the public sector: Evidence from the labor market for teachers*. (Policy research working paper WPS8050). New York, NY: The World Bank.

Bellwether Education Partners (2018). *An independent review of ESSA state plans: Pennsylvania.* Retrived from https://bellwethereducation.org/sites/default/files/Bellwether_ESSA_PlanReview_PA_Final.pdf

Bleiberg, J., & Harbatkin, E. (2018). Teacher evaluation reform: A convergence of federal and local forces. *Educational Policy,* 1-35.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231-268.

Brandt, E. (Apr 2018). Pennsylvania leads nation in teacher strikes, but statewide strike unlikely. *Daily Local News.*

Brunner, E., Hyman, J., & Ju, A. (2019). School finance reforms, teachers' unions, and the allocation of school resources. *Review of Economics and Statistics*, 1-47.

Campbell, A., Converse, P., Miller, W., & Stokes, D. (1980). *The American voter.* Chicago, IL: University of Chicago Press, Midway Reprints.

Campbell, A. F. (2019, Feb 11). Denver teachers are striking for the first time in 25 years. *Vox.*

Candelaria, C. A., & Shores, K. A. (2019). Court-ordered finance reforms in the adequacy era: Heterogeneous causal effects and sensitivity. *Education Finance and Policy*, *14*(1), 31-60.

Casey, L. (2006). The educational value of democratic voice: A defense of collective bargaining in American education. In J. Hannaway & A. J. Rotherham (Eds.), *Collective bargaining in education: Negotiating change in today's schools* (pp. 181-202). Cambridge, MA: Harvard Education Press.

The Center on Standards, Alignment, Instruction, and Learning (C-SAIL) (2019). *State accountability systems and the policies that support them.*

Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science, 24*(2), 140-149.

Cheng, A., Henderson, M. B., Peterson, P. E., & West, M. R. (2018). Public support climbs for teacher pay, school expenditures, charter schools, and universal vouchers**.** *Education Next, 19*(1), 1-26.

Chiang, H., Speroni, C., Herrmann, M., Hallgren, K., Burkander, P., Wellington, A. (2017). *Evaluation of the Teacher Incentive Fund: Final report on Implementation and impacts of pay-for-performance after two years. NCEE 2018-4044.* Washington, D.C.: National Center for Education Evaluation and Regional Assistance.

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, *30*(3), 449-465.

Chingos, M. M., & Whitehurst, G. J. R. (2011). *Class size: What research says and what it means for state policy.* Washington, D.C.: Brookings.

Chingos, M. M. (2012). The impact of a universal class-size reduction policy: Evidence from Florida's statewide mandate. *Economics of Education Review*, *31*(5), 543-562.

Chong, D., & Druckman, J. N. (2007). Framing public opinion in competitive democracies. *American Political Science Review*, *101*(4), 637-655.

Chubb, J. E., & Moe, T. M. (1991). *Politics, markets and America's schools*. Washington, D.C.: The Brookings Institution.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Cohen, D. K., Spillane, J. P., & Peurach, D. P. (2018). The dilemmas of educational reform. *Educational Researcher, 47*(3), 204-212.

Cohen, P. (2016, Feb 21). A rising call to promote STEM education and cut liberal arts funding. *The New York Times.*

Collins, J. E. (2019). Do teachers want democracy? Deliberative culture and teachers' evaluations of schools. *Urban Affairs Review.*

Conley, D. T. (2011). Building on the Common Core. *Educational Leadership*, *68*(6), 16-20.

Cowen, J. M. (2009). Teacher unions and teacher compensation: New evidence for the impact of bargaining. *Journal of Education Finance*, *35*(2), 172-193.

Cowen, J. W., & Strunk, K. O. (2015). The impact of teachers' unions on educational outcomes: What we know and what we need to learn. *Economics of Education Review, 48,* 208-223.

Cowen, J., & Goldhaber, D. (2018). Do bonuses affect teacher staffing and student achievement in high poverty schools? Evidence from an incentive for National Board Certified teachers in Washington State. *Economics of Education Review, 65*, 138-152.

Dahill-Brown, S. E. (2019). *Education, equity, and the states: How variations in state governance make or break reform.* Cambridge, MA: Harvard Education Press.

Darling-Hammond, L. (2003). Keeping good teachers: Why it matters, what leaders can do. *Educational Leadership, 60*(8), 6-13.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13(42),* 1-52.

Day, C., & Gu, Q. (2009). Veteran teachers: Commitment, resilience, and quality retention. *Teachers and Teaching: Theory and Practice, 15*(4), 441-457.

de Ree, J., Muralidharan, K., Pradhan, M., & Rogers, H. (2018). Double for nothing? experimental evidence on an unconditional teacher salary increase in Indonesia. *The Quarterly Journal of Economics, 133*(2), 993-1039.

Dee, T. S., & Jacob, B. (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.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.

Desimone, L. M., & Pak, K. (2017). Instructional coaching as high-quality professional development. *Theory Into Practice, 56*(1), 3-12.

Dever, R., & Lash, M. (2013). Using common planning time to foster professional learning. *Middle School Journal, 45*(1), 12-17.

DiMaggio, P., & Powell, W. (1983). The Iron Cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147-160.

Dismuke, D. P. (2013). *Terms and conditions of employment in collective bargaining agreements negotiated by school boards and teachers associations in pennsylvania school districts for 2012-2013.* Unpublished dissertation.

Dolton, P., & Marcenaro-Gutierrez, O. D. (2011). If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic Policy, 26*(65), 5-55.

Donaldson, M. L., & Woulfin, S. (2018). From tinkering to going "rogue": How principals use agency when enacting new teacher evaluation systems. *Educational Evaluation and Policy Analysis, 40*(4), 531-556.

Dynarski, M. (2018). Is the high school graduation rate really going up? *Evidence Speaks.* Washington, D.C.: Brookings.

Edgerton, A. K., & Desimone, L. M. (2018). Teacher implementation of college- and career- readiness standards: Links among policy, instruction, challenges and resources. *AERA Open, 4*(5), 1-22.

Edgerton, A. K. (2019a). Learning from standards deviation: Three dimensions for building education policy that lasts. *American Educational Research Journal.*

Edgerton, A. K. (2019b). The essence of ESSA: More control at the district level? *Phi Delta Kappan, 101*(2), 14-17.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2016). Selecting growth measures for use in school evaluation systems: Should proportionality matter? *Educational Policy, 30*(3), 465-500.

Elgart, M. A. (2016). Creating state accountability systems that help schools improve: The learning process for every school, like every child, is a personalized journey of continuous improvement. *Phi Delta Kappan, 98*(1), 26.

Esack, S. (2018 Apr). Is partisan political gridlock coming to your school board? Proposed bill would end cross-filing. *The Morning Call.*

Farber, H. S., Herbst, D., Kuziemko, I., & Naidu, S. (2018). *Unions and inequality over the twentieth century: New evidence from survey data.* (NBER Working paper 24587). Cambridge, MA: National Bureau of Economic Research.

Feng, L. (2014). Teacher placement, mobility, and occupational choices after teaching. *Education Economics, 22*(1), 24-47.

Feng, L., Figlio, D., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics, 103*, 1-17.

Feng, L., & Sass, T. R. (2018). The impact of incentives to recruit and retain teachers in “hard-to-staff” subjects. *Journal of Policy Analysis and Management*, *37*(1), 112-135.

Figlio, D., & Loeb, S. (2011). *School accountability*. In E. A. Haushek, S. Machin, & L. Woessman (Eds), *Handbook of the Economics of Education, Vol 3* (pp. 383-421). North- Holland: Elsevier

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, *91*(5-6), 901-914.

Filges, T., Sonne-Schmidt, C. S., & Nielsen, B. C. V. (2018). Small class sizes for improving student achievement in primary and secondary schools: A systematic review. *Campbell Systematic Reviews*, *10*.

Finan, F., Olken, B. A., & Pande, R. (2017). The personnel economics of the developing state. In E. Duflo & A. Banerjee (Eds.), *Handbook of Economic Field Experiments* (Vol. 2, pp. 467-514). North-Holland: Elsevier.

Finger, L. K. (2017). Vested interests and the diffusion of education policy reform across the states. *Policy Studies Journal*, *46*(2), 378-401.

Fiorina, M. P. (1978). Economic retrospective voting in American national elections: A micro-analysis. *American Journal of Political Science*, 426-443.

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, *32*(2), 100-107.

Flavin, P., & Hartney, M. T. (2015). When government subsidizes its own: Collective bargaining laws as agents of political mobilization. *American Journal of Political Science*, *59*(4), 896-911.

Foucault, M., & Sheridan, A. (1975). *Discipline and punish: The birth of the prison.* New York, NY: Random House.

Fullan, M. (2016). The elusive nature of whole system improvement in education. *Journal of Educational Change*, *17*(4), 539-544.

Fullan, M. (2012). *Change forces: Probing the depths of educational reform*. New York, NY: Routledge.

Gadarian, K. S., Goodman, S. W., & Pepinsky, T. B. (2020). *Partisanship, health, behavior, and policy attitudes in the early stages of the COVID-19 pandemic*. Available at SSRN: https://ssrn.com/abstract=3562796

Gamson, D.A. (2015). Differences in common. In J. A. Supovitiz & J. P. Spillane (Eds.), *Challenging standards: Navigating conflict and building capacity in the ear of the Common Core* (pp. 133-144). Lanhan, MD: Rowan & Littlefield.

Garda, R. (2011). The politics of education reform: Lessons from New Orleans. *Journal of Law and Education, 40*(1), 57-103.

Gilraine, M., Macartney, H., McMillan, R. (2018). *Education reform in general equilibrium: Evidence from California's Class Size Reduction* (No. w24191). Cambridge, MA: National Bureau of Economic Research.

Glaeser, E. L., & Ponzetto, G. A. (2014). Shrouded costs of government: The political economy of state and local public pensions. *Journal of Public Economics, 116,* 89-105.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*(1), 134-150.

Goldhaber, D., Lavery, L., Theobald, R., D'Entremont, D., & Fang, Y. (2013). Teacher collective bargaining: Assessing the internal validity of partial independence item response measures of contract restrictiveness. *SAGE Open, 3*(2), 1-13.

Goldstein, D. (2019, Jan 30). In 13 years of education reporting, so much has changed. *The New York Times.*

González Canché, M. S. (2018). Nearby college enrollment and geographical skills mismatch: (Re)conceptualizing student out-migration in the American higher education system. *The Journal of Higher Education, 89*(6), 1-43.

Goodson, I., Moore, S. & Hargreaves, A. (2006). Teacher nostalgia and the sustainability of reform: The generation and degeneration of teachers' missions, memory, and meaning. *Educational Administration Quarterly, 42*(1), 42-61.

Graham, K. A. (2017, Nov 15). School Reform Commission votes to disband. *The Philadelphia Inquirer.*

Grissom, J. A. (2010). The determinants of conflict on governing boards in public organizations: The case of California school boards. *Journal of Public Administration Research and Theory, 20*(3), 601–627.

Grissom, J. A. (2014). Is discord detrimental? Using institutional variation to identify the impact of public governing board conflict on outcomes. *Journal of Public Administration Research and Theory, 24*(2), 289–315.

Gutiérrez, R., & Dixon-Román, E. (2011). Beyond gap gazing: How can thinking about education comprehensively help us (re)envision mathematics education? In Atweh, B., Graven, M., Secada, W., Valero, P. (Eds.) *Mapping equity and quality in mathematics education* (pp. 21-34). Dordrecht: Springer.

Halgren, K., James-Burdumy, S., & Perez-Johnson, I. (2014). *State requirements for teacher evaluation policies promoted by Race to the Top*. NCEE Evaluation Brief. Washington, DC: Institute for Education Sciences. Retrieved from https://ies.ed.gov/ncee/pubs/20144016/pdf/20144016.pdf

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141-1177.

Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141-164.

Hanushek, E. A., & Luque, J. (2000). Smaller classes, lower salaries? The effects of class size on teacher labor markets. In S. W. Laine & J. G. Wards (Eds.), *Using what we know: A review of the research on implementing class-size reduction initiatives for state and local policymakers* (pp. 35-51). Oak Brook, IL: North Central Regional Educational Laboratory, 35051.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review, 30*(3), 466-479.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management, 24*(2), 297-327.

Hess, F. M., & Loup, C. (2008). *The leadership limbo: Teacher labor agreements in America's fifty largest school districts.* Washington, D.C: Thomas B. Fordham Institute.

Hill, P. (2006). The costs of collective bargaining agreements and related district policies. In J. Hannaway & A. J. Rotherham (Eds.), *Collective bargaining in education: Negotiating change in today's schools* (pp. 89-109). Cambridge, MA: Harvard Education Press.

Hoxby, C. M. (2000a). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics, 115*(4), 1239-1285.

Hoxby, C. M. (2000b). Does competition among public schools benefit students and taxpayers?. *American Economic Review, 90*(5), 1209-1238.

Ingersoll, R. M. (2003). *Who controls teachers' work?: Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.

Ingersoll, R. M., & Perda, D. (2010). Is the supply of mathematics and science teachers sufficient? *American Educational Research Journal*, *47*(3), 563-594.

Ingersoll, R., & Merrill, L. (2017). A quarter century of changes in the elementary and secondary teaching force: From 1987 to 2012. Statistical Analysis Report. NCES 2017-092. *National Center for Education Statistics.*

Ingersoll, R. M., & Collins, G. J. (2018). Accountability, control, and teachers' work in American schools. In S. J. Zepeda & J. A. Ponticell (Eds.), *The Wiley Handbook of Educational Supervision* (pp. 157-182.). Hoboken, NJ: Wiley-Blackwell.

Ingle, W. K., & Wisman, R. A. (2018). Extending the work of Cowen and Fowles: A historical analysis of Kentucky teacher contracts. *Educational Policy*, *32*(2), 313-333.

Jackson, C. K. (2012). School competition and teacher labor markets: Evidence from charter school entry in North Carolina. *Journal of Public Economics*, *96*(5-6), 431-448.

Jennings, J. L., & Bearak, J. M. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, *43*(8), 381-389.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1-39.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615-631.

Kaufman, J. H., Opfer, V. D., Bongard, M., & Pane, J. D. (2018). *Changes to what teachers know and do in the Common Core era.* American Teacher Panel findings from 2015 to 2017. Santa Monica, CA: RAND Corporation.

Keefe, J. (2018). *Pennsylvania's teachers are undercompensated – a new pension legislation will cut their compensation even more.* Washington, D.C.: Economic Policy Institute.

Kendi, I. (2017). *Stamped from the beginning: The definitive history of racist ideas in America*. New York, NY: Random House.

Kirst, M. W., & Wirt, F. M. (2009). *The political dynamics of American education* (4th ed.). Richmond, CA: McCutchan Pub Corp.

Koski, W. S., & Horng, E. L. (2007). Facilitating the teacher quality gap? Collective bargaining agreements, teacher hiring and transfer rules, and teacher assignment among schools in California. *Education Finance and Policy, 2*(3), 262-300.

Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016). School organizational contexts, teacher turnover, and student achievement: Evidence from panel data. *American Educational Research Journal*, *53*(5), 1411-1449.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, *113*(485), F34-F63.

Ladd, H. F. (2017). No child left behind: A deeply flawed federal policy. *Journal of Policy Analysis and Management*, *36*(2), 461-469.

Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans.* Chicago, IL: The University of Chicago Press.

Lindstrom, N. (2018 April). After 19 months of negotiations, Pittsburgh Public Schools OKs union contracts. *Triblive.com.*

Lovenheim, M., & Willén, A. (2019). The long-run effects of teacher collective bargaining. *American Economic Journal: Economic Policy, 11*(3), 292-324.

McGuinn, P. (2012). Stimulating reform: Race to the Top, competitive grants and the Obama education agenda. *Educational Policy*, *26*(1), 136-159

Macartney, H., & Singleton, J. D. (2018). School boards and student segregation. *Journal of Public Economics*, *164*, 165– 182.

Malkus, N., & Cummings, A. (2018 Nov). What we're getting wrong about credit recovery. *Education Week.*

Marianno, B. D., Kilbride, T., Theobald, R., Strunk, K. O., Cowen, J. M., & Goldhaber, D. (2017). Cut from the same cloth? Comparing urban district CBAs within states and across the United States. *Educational Policy, 32*(3), 334-359.

Marianno, B. D., & Strunk, K. (2018). The bad end of the bargain?: Revisiting the relationship between collective bargaining agreements and student achievement. *Economics of Education Review, 65*, 93-106.

Marianno, B. D., & Strunk, K. (2018). After Janus. *Education Next, 18*(4), 18-25.

Mathis, W. J. (2017). The effectiveness of class size reduction. *Psychosociological Issues in Human Resource Management*, *5*(1), 176.

Mehta, J. (2013). *The allure of order: High hopes, dashed expectations and the troubled quest to remake American schooling*. New York, NY: Oxford University Press.

Meredith, M., & Grissom, J. (2010). *Partisanship in local elections: Regression discontinuity estimates from unconventional school board races.* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.450.6345&rep=rep1&type=pdf

Merkle, J. S., & Phillips, M. A. (2018). The wage impact of teachers' unions: A meta-analysis. *Contemporary Economic Policy*, *36*(1), 93-115.

Merritt, E. G. (2017). Time for teacher learning, planning critical for school reform. *Phi Delta Kappan*, *98*(4), 31-36.

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.

Montarti, E. (2010). *Pay for performance in Pittsburgh Public Schools: Will it pay off?* Pittsburgh, PA: Allegheny Institute for Public Policy.

Moe, T. M. (2009). Collective bargaining and the performance of the public schools. *American Journal of Political Science*, *53*(1), 156-174.

Moe, T. M. (2015). Vested interests and political institutions. *Political Science Quarterly, 130*(2), 277-318.

Moe, T. M., & Wiborg, S. (2017). Introduction. In T. M. Moe & S. Wiborg (Eds.), *The comparative politics of education: Teachers' unions and education systems around the world* (pp. 1-21). Cambridge, UK: Cambridge University Press.

Moe, T. M. (2017). Teachers' unions in the United States. In T. M. Moe & S. Wiborg (Eds,), *The comparative politics of education: Teachers' unions and education systems around the world* (pp. 22-55). Cambridge, UK: Cambridge University Press.

Morel, D. (2018). *Takeover: Race, education, and American democracy*. New York, NY: Oxford University Press.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, *119*(1), 39-77.

Murnane, R., Singer, J. D., Kemple, J., & Olsen, R. (2009). *Who will teach?: Policies that matter.* Camrbidge, MA: Harvard University Press.

National Council on Teacher Quality. (2012, June). TR3 trends: Teacher planning time. *NCTQ Teacher Trendline.*

National Education Association (2016). *Rankings of the states 2016 and estimates of school statistics 2017.* Washington, D.C.: NEA Research.

OpenDataPA (2018). *Pa school districts boundaries 2017.* Retrieved from https://data.pa.gov/Geospatial-Data/Pa-School-Districts-Boundaries-2017/s629-r52w

Paglayan, A. S. (2018). Public-sector unions and the size of government. *American Journal of Political Science*, *63*(1), 21-36.

Pak, K., Edgerton, A. K., Desimone, L. M., & Song, M. (2019). *Implementation of college- and career-readiness standards: A 50-state comparison.* Unpublished manuscript.

Papay, J. P., & Kraft, M. A. (2016). The myth of the performance plateau. *Educational Leadership*, *73*(8), 36-42.

Papay, J. P., & Hannon, M. (2018). *The effects of school turnaround strategies in Massachusetts.* Paper presented at the Association for Public Policy Analysis and Management. Washington, D.C.

Pennsylvania Department of Education (2018). *Keystone exams.*

Pennsylvania Department of State (2018). *PA full voter export.*

Pennsylvania State Education Association (2018). *Collective bargaining.*

Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *The Journal of the Association for Public Policy Analysis and Management*, *26*(4), 909-950.

Policy Analysis for California Education (PACE) (2018). PACE and USC Rossier polls. Retrieved from http://edpolicyinca.org/polls

Perrillo, J. (2004). Beyond "Progressive" reform: Bodies, discipline, and the construction of the "professional teacher" in Interwar America. *History of Education Quarterly*, *44*(3), 337-363.

Ravitch, D. (2006-2007, Winter). Why teacher unions are good for teachers—And the

public. *American Educator.* Retrieved from http://www.aft.org/periodical/american-educator/winter-2006-2007/why-teacher-unions-are-good-teachers-and

Reardon, S. F., & Raudenbush, S. W. (2006). A partial independence item response model for surveys with filter questions. *Sociological Methodology*, *36*(1), 257-300.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G.J. Duncan & R.J. Murname (Eds.), *Whither opportunity* (pp. 91-116). New York, NY: Russell Sage Foundation.

Reeves, P. M., Pun, W. H., & Chung, K. S. (2017). Influence of teacher collaboration on job satisfaction and student achievement. *Teaching and Teacher Education*, *67*, 227-236.

Research for Action (2018). *Pennsylvania School Data Project.* Retrieved from https://www.researchforaction.org/pa-school-data-project/

Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., & Griffin, B. A. (2014). *Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the Twang Package* (R vignette). Santa Monica, CA: RAND.

Rincke, J. (2006). Competition in the public school sector: Evidence on strategic interaction among US school districts. *Journal of Urban Economics*, *59*(3), 352-369.

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.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Methods Research*, *3*(3), 208-222.

Sasko, C. (2017 Jun). Teachers approve contract with Philly schools. *Phillymag.com*

Schneider, J. (2017). *Beyond test scores: A better way to measure school quality*. Cambridge, MA: Harvard University Press.

Schueler, B. E., Goodman, J. S., & Deming, D. J. (2017). Can states take over and turn around school districts? Evidence from Lawrence, Massachusetts. *Educational Evaluation and Policy Analysis*, *39*(2), 311-332.

Schueler, B. E. (2018). *School district turnaround: Learning from leadership in Lawrence, Massachusetts.* Policy Brief. Cambridge, MA: Harvard Kennedy School Rappaport Institute for Greater Boston.

Schworm, P. (2012, March 23). Lawrence ex-school chief gets jail term. *The Boston Globe.*

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.

Scribner, C. F. (2015). Beyond the metropolis: The forgotten history of small-town teachers' unions. *American Journal of Education*, *121*(4), 531-561.

Shi, Y., & Singleton, J. D. (2018). Expertise and independence on governing boards:

Evidence from school districts (CEPA Working Paper No.18-21). Palo Alto, CA: Stanford Center for Education Policy Analysis.

Soja, E. W. (2013). *Seeking spatial justice*. Minneapolis, MN: University of Minnesota Press.

Springer, M. G., Ballou, D., Hamilton, L., Le, V. N., Lockwood, J. R., McCaffrey, D. F., Pepper, M., & Stecher, B. M. (2011). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching (POINT).* Washington, D.C.: Society for Research on Educational Effectiveness.

Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, *27*(5), 556-563.

Startz, D. (2016). *Teacher pay around the world.* Washington, D.C.: Brookings.

Stecher, B. M., Holtzman, D. J., Garet, M. S., Hamilton, L. S., Engberg, J., Steiner, E. D., Robyn, A., Baird, M. D., Gutierrez, I. A., Peet, E. D., de los Reyes, I. B., Fronberg, K., Weinberger, G., Hunter, G. P., & Chambers, J. (2018). *Improving teaching effectiveness: Final report: The Intensive Partnerships for Effective Teaching through 2015–2016.* RR-2242-BMGF. Santa Monica, CA: RAND Corporation, RR-2242-BMGF, 2018.

Steinberg, M. P., & Quinn, R. (2015). A tale of two decades: New evidence on adequacy and equity in Pennsylvania. *Journal of Education Finance*, *40*(3), 273-299.

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy, 10*(4), 535-572.

Steinberg, M. P., & Kraft, M. A. (2017). The sensitivity of teacher performance ratings to the design of teacher evaluation systems. *Educational Researcher, 46*(7), 378-396.

Strauss, V. (2016, Aug). The "seven deadly sins" of Common Core – by an English teacher. *The Washington Post.*

Strunk, K. O., & Grissom, J. A. (2010). Do strong unions shape district policies? Collective bargaining, teacher contract restrictiveness, and the political power of teachers' unions. *Educational Evaluation and Policy Analysis, 32*(3), 389-406.

Strunk, K. O., & Reardon, S. F. (2010). Measuring the strength of teachers' unions: An empirical application of the partial independence item response approach. *Journal of Educational and Behavioral Statistics, 35*(6), 629-670.

Strunk, K. O., & McEachin, A. (2011). Accountability under constraint: The relationship between collective bargaining agreements and California schools' and districts' performance under No Child Left Behind. *American Educational Research Journal, 48*(4), 871–903.

Strunk, K. O., Cowen, J. M., Goldhaber, D., Marianno, B. D., Kilbride, T., & Theobald, R. (2018). It is in the contract: How the policies set in teachers' unions' collective bargaining agreements vary across states and districts. *Educational Policy, 32*(2), 280-312.

Smith, M., & Davey, M. (2019 Oct 31). Chicago teachers' strike, longest in decades, ends. *The New York Times.*

Stein, P. (2018 Apr 27). D.C. schools could face sharp dip in graduation rate after scandal. *The Washington Post.*

Sunderman, G. L., & Kim, J. S. (2007). The expansion of federal power and the politics of implementing the No Child Left Behind Act. *Teachers College Record, 109*(5), 1057-1085.

Supovitz, J. A. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement.* Cambridge, MA: Harvard Education Press.

Supovitz, J. (2017). Social media is the new player in the politics of education. *Phi Delta Kappan, 99*(3), 50-55.

Swain, W. A., Rodriguez, L. A., & Springer, M. G. (2019). Selective retention bonuses for highly effective teachers in high poverty schools: Evidence from Tennessee. *Economics of Education Review, 68*, 148-160.

Taylor, L. L. (2006). Comparable wages, inflation, and school finance equity. *Education Finance and Policy, 1*(3), 349-371.

Tetlock, P. E., Vieider, F. M., Patil, S. V., & Grant, A. M. (2013). Accountability and ideology: When left looks right and right looks left. *Organizational Behavior and Human Decision Processes, 122*(1), 22-35.

Trujillo, T., Scott, J., & Rivera, M. (2017). Follow the yellow brick road: Teach for America and the making of educational leaders. *American Journal of Education, 123*(3), 353-391.

Viteritti, J. P. (2012). Educational innovation and the law: The federal role in school reform: Obama's "Race to the Top." *Notre Dame Law Review, 87*, 2087–2261.

Wei, X. (2012). Are more stringent NCLB state accountability systems associated with better student outcomes? An analysis of NAEP results across states. *Educational Policy, 26*(2), 268-308.

White, R. S. (2018). Who do state policy makers listen to? The case of teacher evaluation. *Phi Delta Kappan, 99*(8), 13-18.

Winter, J. V. (2011). Teacher salaries and teachers unions: A spatial econometric approach. *IRL Review, 64*(4), 747-764.

Woessmann, L. (2011). Cross-country evidence on teacher performance pay. *Economics of Education Review, 30*(3), 404-418.

Woulfin, S. L., Donaldson, M. L., & Gonzales, R. (2016). District leaders' framing of educator evaluation policy. *Educational Administration Quarterly, 52*(1), 110-143.

Wrabel, S. L., Saultz, A., Polikoff, M. S., McEachin, A., & Duque, M. (2018). The politics of elementary and secondary education act waivers. *Educational Policy, 32*(1), 117-140.

Zinth, K. (2009). *Maximum P-12 class size policies.* Denver, CO: Education Commission of the States.

Zwerling, H. L. (2008). Pennsylvania teachers strikes and academic performance. *Journal of Collective Negotiations, 32*(2), 151-172.