Department of Economics, University of Warwick

Monash Business School, Monash University

as part of

Monash Warwick Alliance

The Educational Impact of Weakening Teachers' Unions: Evidence from Wisconsin

Jack Kemp

Warwick-Monash Economics Student Papers

September 2022 No: 2022-39

ISSN 2754-3129 (Online)

The Warwick Monash Economics Student Papers (WM-ESP) gather the best Undergraduate and Masters dissertations by Economics students from the University of Warwick and Monash University. This bi-annual paper series showcases research undertaken by our students on a varied range of topics. Papers range in length from 5,000 to 8,000 words depending on whether the student is an undergraduate or postgraduate, and the university they attend. The papers included in the series are carefully selected based on their quality and originality. WM-ESP aims to disseminate research in Economics as well as acknowledge the students for their exemplary work, contributing to the research environment in both departments.

"*We are very happy to introduce the Warwick Monash Economics Student Papers (WM-ESP). The Department of Economics of the University of Warwick and the Economics Department at Monash University are very proud of their long history of collaboration with international partner universities, and the Monash Warwick Alliance reflects the belief in both Universities that the future will rely on strong links between peer Universities, reflected in faculty, student, and research linkages. This paper series reflects the first step in allowing our Undergraduate, Honours, and Masters students to learn from and interact with peers within the Alliance.*"

Ben Lockwood (Head of the Department of Economics, University of Warwick) and Michael Ward (Head of the Department of Economics, Monash University)

Recommended citation: Kemp, J. (2022). The Educational Impact of Weakening Teachers' Unions: Evidence from Wisconsin. *Warwick Monash Economics Student Papers* 2022/39.

[1] Warwick Economics would like to thank Lory Barile, Gianna Boero, and Caroline Elliott for their contributions towards the selection process.

The Educational Impact of Weakening Teachers' Unions: Evidence from Wisconsin

Jack Kemp[*]

28th April 2022

Abstract

The impact of teachers' unions on educational outcomes is theoretically and empirically ambiguous. This study aims to provide additional evidence to this debate by analyzing the educational impact of Wisconsin's Act 10, which significantly weakened teachers' unions. A unique school-level dataset is constructed containing data on educational outcomes, high-school and district-level characteristics in Wisconsin and Illinois. Schools in treated and control groups are matched using propensity score matching. A difference-in-difference methodology is used to mitigate problems surrounding simultaneity and omitted variable bias from standard OLS regression. This paper uses an event study design and a triple difference model to estimate the effect of Act 10 on the dropout rate and subject proficiency rate. The triple difference model used has lower power but is more robust to the parallel trends assumption as it takes into account differences in economic conditions between states. Models suggest that weakening teachers' unions worsens educational outcomes, providing support for the union-voice hypothesis.

Keywords: Education, Teacher unions, Event study, Triple difference, Propensity score matching

JEL Classifications: I20; I28; J45; J50

[*] Email: jack.kemp45@gmail.com. The online appendix and the associated files can be found on GitHub at: https://github.com/jackkemp314/Undergraduate-dissertaion-appendix. I would like to express my gratitude to my supervisor Natalia Zinovyeva for providing me with valuable support and to my parents who sparked my interest in studying education and the impact of teachers' unions.

1. Introduction

To reconcile state budgets after the Great Recession, American policymakers implemented anti-union legislation curbing the influence of public-sector unions. These laws allowed states to reduce the salaries and pensions of public-sector workers and generate considerable cost savings (Freeman and Han, 2012), motivated by the political zeitgeist that “government costs more but does less” (DiSalvo, 2015). Around 60% of teachers were covered by a collectively bargained contract (Frandsen, 2016) until 2011 when Idaho, Indiana, Tennessee, and Wisconsin passed legislation that limited the extent of teacher bargaining. Wisconsin’s Budget Repair Bill, known as Act 10, is the most well-documented of these antiunion laws. Act 10 substantially limited the scope of teacher collective bargaining and union fundraising activities and was seen as the most significant attack on labour rights in the US since the 1980’s (Kaufman, 2021). Figure 1 shows the impact of Act 10 on public-sector union membership in Wisconsin. Membership fell from 50% to 37% in the first year of the policy and continued to fall to below the US average. The neighboring state Illinois was unaffected. This figure shows that Act 10 had a substantial impact on public-sector union power in Wisconsin.

The direction of the impact of unions on educational outcomes is theoretically and empirically ambiguous. This study aims to provide additional evidence on this debate by analyzing the impact of Act 10 on dropout and subject proficiency rates. Act 10 offers an interesting natural experiment as it represents an exogenous shock to union power in Wisconsin, impacting public schools in Wisconsin but not charter schools or schools in the neighbouring state Illinois. Results for the subject proficiency rate are compared to Baron (2018), who uses within-state variation in collective bargaining agreements to find that Act 10 negatively affected test scores in the short-run.

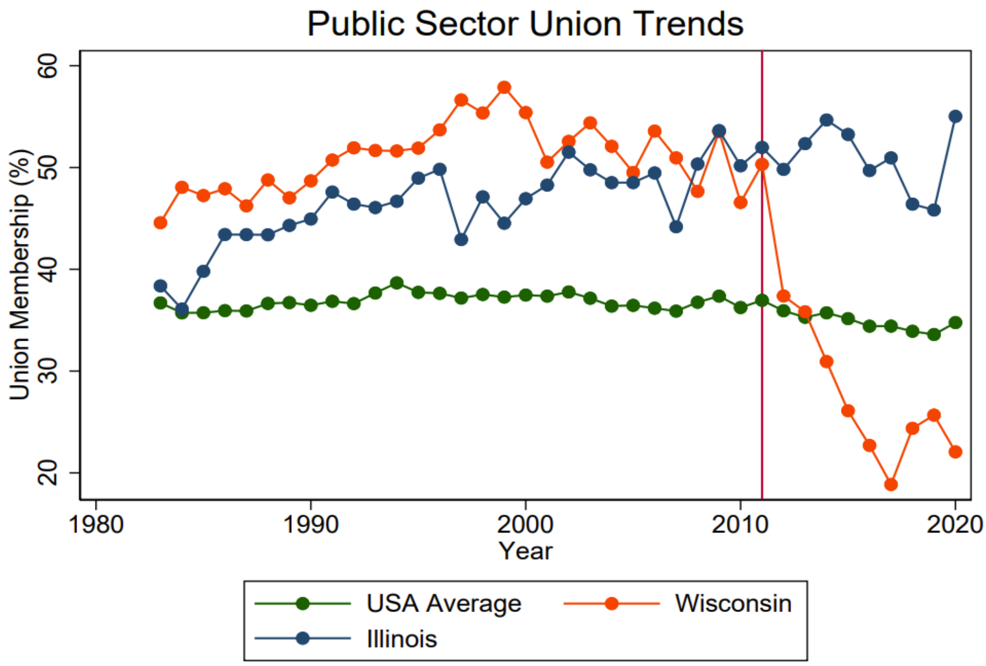

Figure 1: Public-sector union membership rates over time

This study employs an event study design and triple difference model to estimate the effect of Act 10 on educational outcomes. The necessary identification assumption for the event study design is one of parallel trends, which requires that public schools in Wisconsin would have followed the same trend in educational outcomes as public schools in Illinois if Act 10 never occurred. The triple difference model controls for economic differences between states and is more robust to this parallel trends assumption, relying only on parallel trends in relative outcomes for public and charter schools in Wisconsin and Illinois. Results are robust to matching and are not driven by changes in student composition.

There are two main ways that this research contributes to the literature. Firstly, this study extends the analysis to the dropout rate, whereas past literature focuses on test scores. This paper also gives more robust estimation than Baron (2018). It uses a triple difference design which controls for both within-state spillover effects and differences in economic conditions between states and uses propensity score matching to generate a more valid control group.

2. Literature review

There is extensive research on the private and social returns to education. Empirical studies find strong correlations between education and future incomes (Chetty et al., 2011) in accordance with Human Capital Theory (Schultz, 1961; Mincer, 1974; Becker, 1993). Higher levels of education are associated with improved health and life satisfaction (Cutler and Lleras-Muney, 2006; Grossman, 2006) and schooling can affect preferences by making individuals more patient and goal oriented and less likely to engage in risky behaviour (Oreopoulos and Salvanes, 2011). Importantly, from a policy perspective, the social returns to education are greater than the private returns. Education is one of the most important predictors for social engagement and trust (Easterbrook et al., 2016; Helliwell and Putnam, 2007) and is associated with lower levels of criminality (Lochner and Moretti, 2004). Rauch (1993) estimates that knowledge externalities (the diffusion and growth of knowledge in society) increase the returns to education by 3-5 percentage points. Therefore, Act 10 has the potential to have a large and lasting economic impact and it is important to study the educational effects of this policy.

The societal impact of unions is well researched. Economic theory assumes that unions aim to maximise the utility of their members (Oswald, 1985) by influencing the real wage, employment levels, and other factors such as job conditions. Empirical studies find that unions lead to higher wages and job security for their members (Hammer and Avgar, 2005; Meng, 1990) and public-sector unions have been found to generate positive wage spillovers for non-members (Rosenfeld and Denice, 2019). On the other hand, unions negatively impact the employment level in an industry (Wooden and Hawke, 2000) but their effect on labour productivity can vary by country (Doucouliagos and Laroche, 2003). In their seminal paper, Freeman and Medoff (1985) find that unions reduce economic disparities via a more compressed wage distribution. Card, Lemieux and Riddell (2003) find that the observed decline in union power can explain 14% of the growth in wage inequality for men in the US between 1973-2001.

There are two contrasting theories on the educational impact of teachers' unions. The rent-seeking model suggests that teachers' unions can worsen student outcomes by reallocating resources towards teachers and away from other inputs in the education production function (Hoxby, 1996), distorting the optimal allocation of educational inputs. Also, by protecting teachers from being fired, unions can lower average teacher quality and reduce teacher effort. On the other hand, the union-voice hypothesis suggests that giving teachers the ability to influence their working environment can increase their productivity (Freeman 1980; Gunderson 2005) and better working conditions could encourage higher ability individuals to select into teaching. This is supported by the fact that after Act 10, the number of incoming Wisconsin college freshmen who intended to become teachers dropped by a third between 2012 and 2018 (Wisconsin Policy Forum, 2021).

Empirical research does not provide conclusive evidence on the educational impact of teachers' unions. Lovenheim and Willén (2019) use a multistate difference-in-difference framework to suggest that exposure to teacher collective bargaining laws worsens test scores and future incomes, especially for men. Studies focusing specifically on the impact of Act 10 on test scores vary depending on the empirical strategy used, suggesting a need for more robust estimation techniques. Baron (2018) uses within-state variation in collective bargaining agreements to find that Act 10 reduced test scores by 22% of a standard deviation, with this effect largely driven by declines in the lower half of the student performance distribution. However, this study could be biased by within-state spillover effects. Roth (2019) analyses the impact of the increase in teacher retirements following Act 10 using the fraction of teachers over the minimum retirement age of 55 in 2011 as an instrument for the fraction of teachers retiring in a school-grade level. The author finds that test scores improved in grades whose teachers retired following the reform. This instrumental variable (IV) strategy relies on the assumption that school-grade levels with higher fractions of teachers over the minimum retirement age were no more likely to experience administrative changes or other shocks that influenced educational quality. Note that IVs estimate the local average treatment effect which may not be the same as the average treatment effect. The effect of Act 10 on different districts is likely heterogeneous. Certain districts implemented performance related pay following Act 10, which improved teacher quality (Biasi, 2018) and longer-term test scores for these districts (Baron, 2021) and subsequently widened within-state educational inequality. These results suggest that on aggregate short-term test scores may have fallen, but certain schools did well out of Act 10 due to their ability and willingness to implement performance related pay.

The literature is lacking in its analysis of educational outcomes other than test scores. To the author's knowledge, this study will be the first to analyse the impact of Act 10 on dropout rates. This allows a broader view of the educational impact of Act 10 since the dropout rate is an important metric to consider in school performance (Rumberger and Palardy, 2005). A school provides a variety of essential functions for child development such as socialisation, support for vulnerable children, and development of life skills (Lewit, 1992), but dropouts do not take full advantage of these resources. Data from the US Bureau of the Census (1991) shows that high-school dropouts were twice as likely to have incomes below the federal poverty level compared to high-school graduates.

3. Empirical Strategy

Simple OLS regressions of educational outcomes on a measure of union strength would likely lead to biased results due to omitted variable or simultaneity bias. For example, areas where teachers are treated particularly poorly may have greater union power and experience worse educational outcomes, making causal inference difficult. To address this issue, this study will use a difference-in-difference methodology exploiting variations in exposure to Act 10.

3.1 Difference-in-Difference Models

A difference-in-difference methodology is used to identify the effect of Act 10 on educational outcomes. Models are weighted by school size and standard errors are clustered at the school level; otherwise, standard errors will be biased downwards due to serial correlation (Bertrand, Duflo, and Mullainathan, 2004).

The event study design compares public schools in Wisconsin (treated) to public schools in Illinois (control) since Illinois is not directly affected by Act 10. Neighbouring states are a natural control group to consider because it is likely that they share similar characteristics (climate, social norms etc.) and is the strategy employed by Lovenheim and Willén (2019). This model is informative for two reasons. Firstly, the model can provide evidence for the parallel trends assumption by testing if pre-trends are insignificant ($\beta_j = 0, \forall j < 0$). Secondly, the model allows for dynamic treatment effects so can show how the treatment effect evolves over time.

School-level fixed effects are included to account for unobservable time-invariant school differences. This is important: Bau and Das (2020) find that observed teacher characteristics account for less than 5% of teacher value added (how much a teacher improves test scores for a class), so unobserved characteristics are important to control for. Time fixed effects are included to control for time varying factors that affect all schools, such as general economic conditions. The event study design takes the general form:

$$Y_{its} = \theta_i + \delta_t + \sum_{j \neq -1} \beta_j 1\{s = j\} + X_{its}\gamma + \varepsilon_{its} \tag{1}$$

With Y_{its} as the educational outcome of interest for school i at time t and event time s (s = 0 when treatment occurs in 2011). θ_i and δ_t are school and time fixed effects respectively, and X represents a vector of covariates used to increase the statistical power of the estimates. The β coefficients estimate the dynamic treatment effect. Event time -1 is excluded due to collinearity.

The necessary identification assumption is one of parallel trends, meaning treated schools are assumed to follow the same path as untreated schools in the counterfactual that treatment never occurs. There are a few strategies one can use to increase confidence in this assumption. Firstly, figures 4 and 5 show raw plots of the key outcome variables. Visual inspection of pre-trends shows that the dropout rate followed similar trends for Wisconsin and Illinois before 2011, increasing our confidence in the parallel trends assumption. Trends also appear to be quite similar for the proficiency rate. Figures 7 and 8 show the coefficient plots from the event study design which can be used to check for pre-emptive behaviour. If lags are insignificant, this increases our confidence in the parallel trends assumption. Most of the lags for the dropout rate are insignificant. However, two of the four lags are significant for the proficiency rate, so we cannot be fully confident in the parallel trends assumption.

The triple difference model is more robust to the parallel trends assumption and is used to control for economic differences between states. The estimate is computed as the difference between two difference-in-difference estimators (Olden and Moen, 2020). For intuition, firstly, we could compare charter schools to public schools in Wisconsin, since charter schools are not affected directly by the law. However, there could be within-state spillover effects that bias results. For example, worsening job conditions in public schools following Act 10 could encourage high-quality public-school teachers to switch to charter schools and improve relative charter school performance. Hensvik (2012) finds wage spillover effects for public schools when studying the entry of private schools in Sweden. Secondly, we could compare public schools in Wisconsin to public schools in Illinois, which is the strategy used in the event study design. However, these states might be affected by different economic conditions that cause trends in educational outcomes to differ. The triple difference estimator differences these two differences to remove the respective biases and identifies the causal effect if general economic differences between states does not affect the relative outcomes of charter and public schools. The triple difference model takes the form:

$$\begin{aligned} Y_{it} = {} & \alpha + \beta_1 State_i + \beta_2 Charter_i + \beta_3 Post_t + \beta_4 State_i * Charter_i + \beta_5 State_i * Post_t + \\ & \beta_6 Charter_i * Post_t + \delta State_i * Charter_i * Post_t + \varepsilon_{it} \end{aligned} \tag{2}$$

The three-way interaction term δ represents the triple difference estimator. As a caveat, around 4.88% of schools are charter schools in this dataset so will likely be a low powered model.

3.2 Propensity Score Matching

Propensity score matching (Rosenbaum and Rubin, 1983) is a powerful tool for causal inference, particularly in conjunction with a difference-in-difference framework. A propensity score is the probability of being assigned the treated group conditional on observed characteristics. In this setting, the propensity score is the probability a school is treated by Act 10, given school and district-level characteristics. Propensity score matching allows one to design a non-randomized study so that it replicates aspects of a randomized control trial (Austin, 2011) since the observed baseline characteristics should be similar between groups. A concern with difference-in-difference models is that treated and control groups may differ in factors related to their trends over time (Stuart et al., 2014). Propensity score matching offers a solution: in this study, schools are matched on their propensity score and schools are only kept if they have common support (figure 2).

This is a more robust estimation method since propensity score matching for difference-in-difference models can reduce model dependence and increase robustness to misspecification (Imai and Kim, 2021). However, there is a trade-off between the strictness of the matching criteria and statistical power. Whilst stricter matching may lead to comparison of more similar observations, it also will reduce the number of observations in the sample. Since there are not a vast number of observations to start with, a relatively non-restrictive distance criterion is used to maintain the sample size and remove only the least similar observations. Specifically, observations are one-to-one nearest neighbour matched using Mahalanobis distance as the distance criterion. A probit model was used for matching.

428 out of 4982 observations were off support so were dropped from the matched dataset. Figure 3 shows a histogram plotting the propensity score by treatment status. There is some overlap in the propensity score between treated and control groups, although there are still differences at the extremes. This suggests that high schools in Wisconsin and Illinois have significantly different characteristics and simply comparing differences in means in educational outcomes will not identify the causal effect. However, a difference-in-difference strategy is still valid if the parallel trends assumption holds.

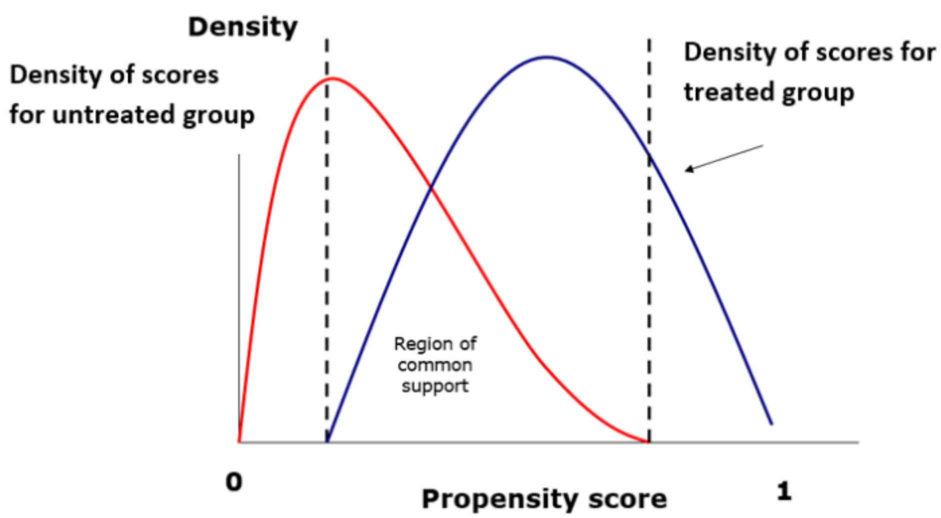

Figure 2: Common support for propensity score matching

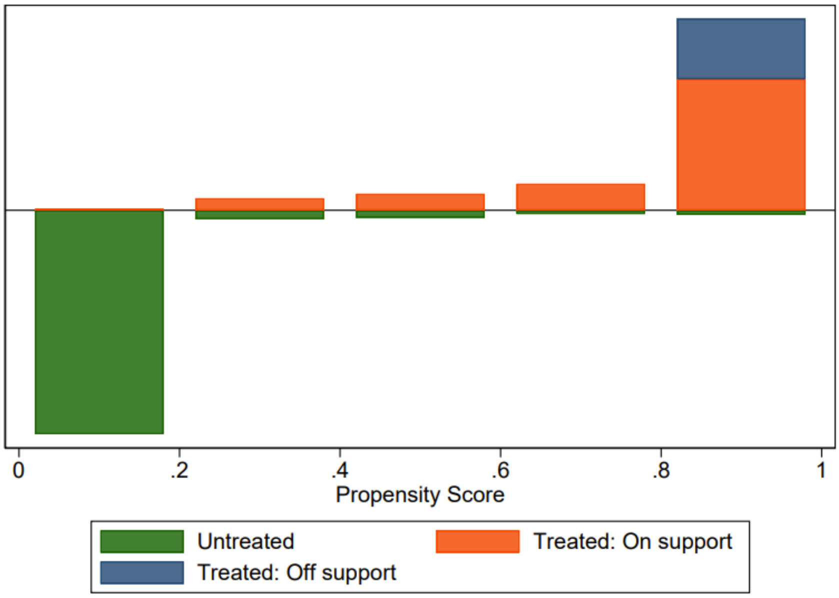

Figure 3: Histogram plotting propensity score by treatment status

4. Data

This study is based on a sample of high schools from Wisconsin and Illinois over the period 2006-2015. A unique school-level dataset is constructed using data from Wisconsin's Department of Public Instruction and the Illinois State Board of Education. In 2006, 324 high schools from Wisconsin and 629 high schools from Illinois were included in the dataset (see table 2).

Table A.1 in the online appendix gives a description of variables included in the dataset. There are three primary outcome variables included which measure different aspects of school performance. The subject proficiency rate measures the percentage of students who achieved a grade of proficient or above in composite math and ELA (English Language Arts) tests. Test scores are commonly used to assess the educational impact of a policy, although normalized test scores are often used rather than proficiency rates. The proficiency rate was chosen as an outcome variable due to data limitations; raw or normalized test scores were not available for all years for Wisconsin or Illinois. The definition of the dropout rate can vary by study (Lehr et al., 2004; Thurlow, Sinclair, and Johnson, 2002). In this case, the dropout rate is defined as the percentage of students who drop out in a single year without completing high school. This is known as the event rate. Finally, the truancy rate is defined as the percentage of students absent from school without a legitimate excuse for five or more school days during a semester. It was included in the dataset to give an insight into the effects of Act 10 on student behaviour. Unfortunately, the definition for the truancy rate changes in 2011 for Illinois (see online appendix A.1). Figure 6 shows the truancy rate in Illinois artificially increase in 2011, so this variable cannot be used in the difference-in-difference analysis.

The dataset also contains teacher outcome variables which will be used to suggest potential mechanisms behind the observed trends in educational outcomes (see online appendix). The percentage of teachers with a master's degree can indicate teacher quality in a school, under the assumption that teachers with higher qualifications are better at teaching. Goldhaber and Brewer (2000) find that students taught by a teacher holding a master's degree perform better on tests on average, but only for teachers with subject-specific degrees. Average teaching experience (not school specific) can also be used to indicate teacher quality if more experienced teachers are better at teaching. Papay and Kraft (2015) find evidence that a teacher's value-added increases rapidly early in their careers and continues to increase to the end of their careers as human capital is accumulated. Average teacher salary can affect teacher productivity and their incentive to work hard (Britton and Propper, 2016; Greenwald and Stiglitz, 1988; Mbiti, Romero and Schipper, 2019) and there is evidence that paying teachers more improves student achievement (Hendricks, 2014).

School and district-level characteristics are used as control variables in the specifications to increase the statistical power of estimates. School size is estimated to have a non-linear relationship, in the form of an inverted-U, with school performance (Bradley and Taylor, 2001) but small high schools have been found to be more effective for disadvantaged groups (Stiefel et al., 2000). The racial makeup of the school is important for school performance even after controlling for socioeconomic and geographic factors (Alexander et al., 1994; Carl and Caldas, 1997; Murphy and Zirkel, 2015). Local property taxes and total school district revenue are included to control for differences in funding between districts which can affect school performance (Venteicher, 2005).

The final dataset used for empirical analysis is matched on observable characteristics using propensity score matching and schools are kept if they have common support (12.5% of observations dropped). Tables 1 and 2 show baseline characteristics in 2006 for high schools in Wisconsin and Illinois pre- and post-match. Whilst characteristics converge after matching, there are still significant differences in means for all observable characteristics. In the final dataset, the average Illinois school had 210 more students than the average Wisconsin school and is more racially diverse (74.43% of students in Illinois are white compared to 90.95% in Wisconsin). Teachers in Illinois on average had an extra 2.6 years of experience, are 11.29 percentage points more likely to have a master's degree and earnt $12,157 more per year.

Figures 4-5 show plots of educational outcome variables over time. Throughout the period, the dropout rate for Illinois was around 2 percentage points higher than the dropout rate for Wisconsin. Visually, the dropout rates for Wisconsin and Illinois seem to follow common trends before 2011. Post Act 10, trends diverged: the dropout rate fell 19.71% for Illinois and 14.67% for Wisconsin. The proficiency rate for Illinois is around 23 percentage points higher compared to Wisconsin during the period. Trends do seem to follow a similar path pre-2011, although there appears to be some divergence in 2007-09. Post Act 10, the proficiency rate increased by 5.34% for Wisconsin but fell 1.52% for Illinois.

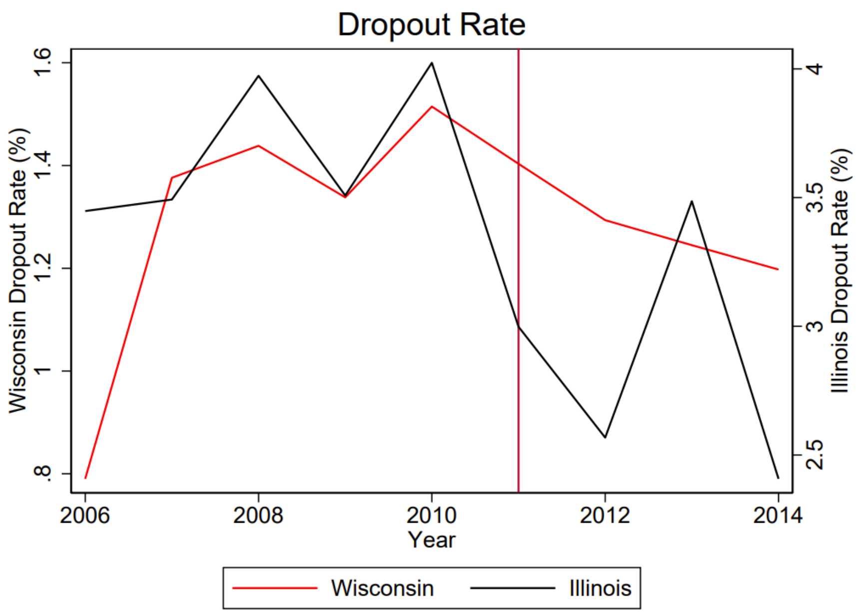

Figure 4: Dropout rate over time

Table 1: Pre-match baseline characteristics

2006 (pre-matching)	(1) Both States	(2) Wisconsin	(3) Illinois	(4) Wisc. - Illinois
School size	816.73	660.86	910.76	-249.9***
	(795.5)	(553.3)	(898.3)	(48.71)
Master's	40.07	32.67	44.52	-11.85***
	(15.28)	(3.413)	(17.74)	(0.878)
Teaching exp	12.73	11.41	13.53	-2.123***
	(2.513)	(1.084)	(2.779)	(0.142)
Salary	48755.70	41170.94	53330.65	-12,160***
	(11602.8)	(2910.5)	(12455.1)	(619.3)
Property taxes	1.9e+08	2.1e+07	3.0e+08	-2.787e+08***
	(521769270.5)	(45511230.8)	(642644185.0)	(3.149e+07)
Total revenue	4.4e+08	7.7e+07	6.7e+08	-5.914e+08***
	(1.19169e+09)	(217904938.1)	(1.46720e+09)	(7.227e+07)
Percent white	76.51	88.91	69.05	19.85***
	(32.67)	(16.67)	(37.35)	(1.936)
Percent black	13.67	4.87	18.01	-13.14***
	(27.32)	(13.38)	(31.14)	(1.759)
N	1111	418	693	

mean coefficients; sd in parentheses
*** p<0.01, ** p<0.05, * p<0.1

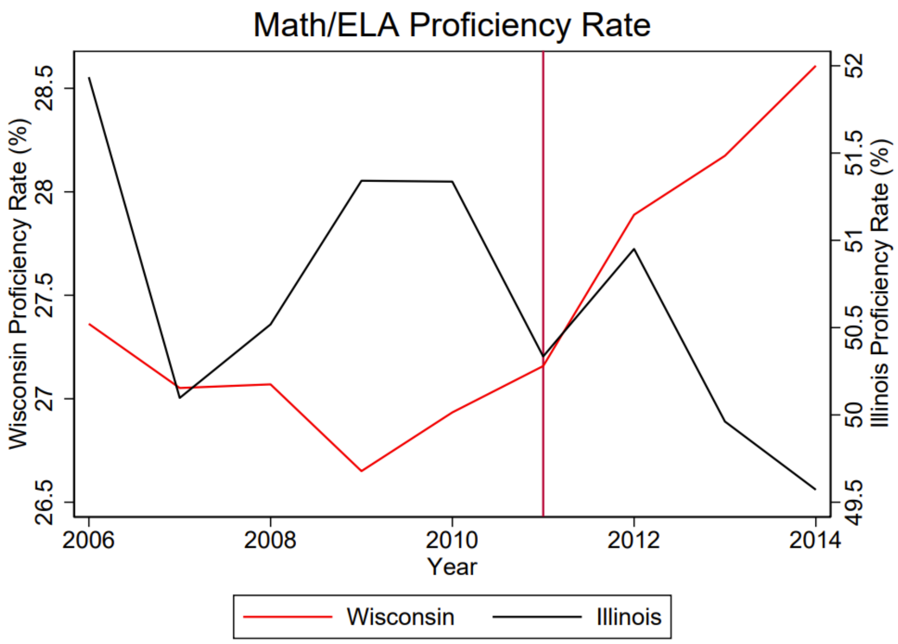

Figure 5: Proficiency rate over time

Table 2: Post-match baseline characteristics

2006 (Post-matching)	(1) Both States	(2) Wisconsin	(3) Illinois	(4) Wisc. - Illinois
School size	882.47	743.74	953.93	-210.2***
	(804.5)	(550.8)	(900.0)	(54.62)
Master's	40.38	32.93	44.22	-11.29***
	(15.61)	(3.306)	(17.90)	(1.003)
Teaching exp	13.05	11.33	13.94	-2.604***
	(2.273)	(1.032)	(2.231)	(0.131)
Salary	49278.06	41254.40	53411.07	-12,157***
	(11791.3)	(2827.3)	(12503.3)	(703.9)
Property taxes	1.7e+08	1.6e+07	2.5e+08	-2.338e+08***
	(493595149.2)	(26675813.3)	(591911509.8)	(3.291e+07)
Total revenue	3.8e+08	4.0e+07	5.5e+08	-5.134e+08***
	(1.12448e+09)	(58698808.2)	(1.35103e+09)	(7.511e+07)
Percent white	80.05	90.95	74.43	16.52***
	(28.94)	(8.910)	(33.70)	(1.906)
Percent black	10.20	2.30	14.26	-11.96***
	(23.41)	(4.507)	(27.78)	(1.554)
N	953	324	629	

mean coefficients; sd in parentheses; propensity score matched with common support
*** p<0.01, ** p<0.05, * p<0.1

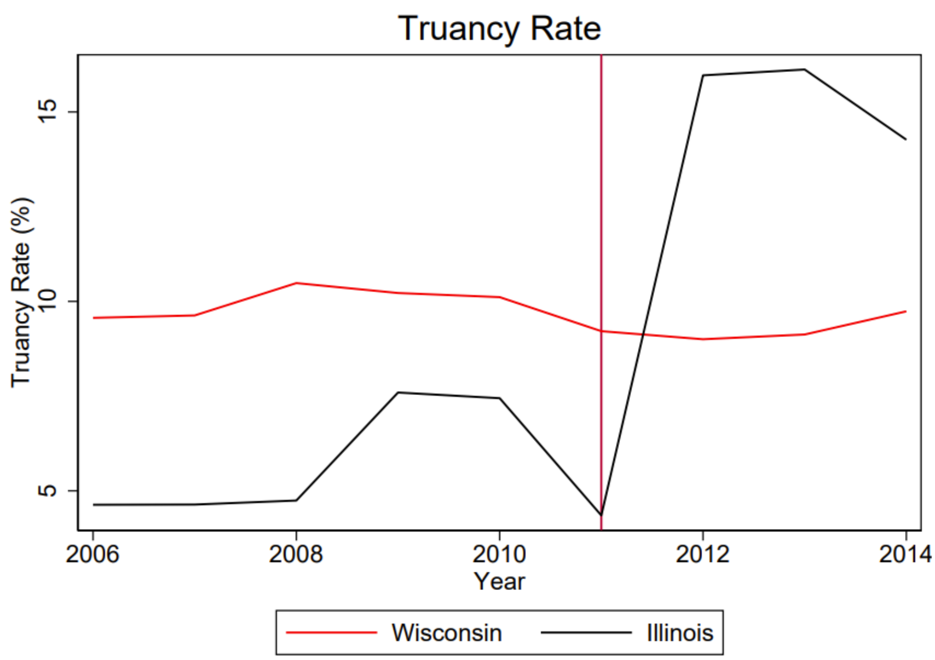

Figure 6: Truancy rate over time

5. Results

5.1 Event Study Design

Results from the event study design are shown in table A.2 in the online appendix which presents coefficient estimates and standard errors leading to and following exposure to Act 10. Figures 7 and 8 provide a visual representation of the estimated coefficients for each event time along with 95% confidence intervals.

For the dropout rate, all but one of the pre-treatment coefficients are statistically insignificant at the 5% level which offers evidence that public schools in Wisconsin and Illinois had a similar evolution in the dropout rate prior to treatment. The treatment effect is statistically significant and appears to be relatively constant over time, suggesting that Act 10 may have a lasting impact on educational outcomes. This model suggests that Act 10 contemporaneously increased the dropout rate for public schools in Wisconsin by 1.188 percentage points, which indicates worsening educational outcomes.

Two of the four lags for the proficiency rate are statistically significant which means we should be cautious in interpreting the results for this variable, and more confidence should be placed in the triple difference model's findings. However, some lags are still insignificant, and the dropout rate appears to satisfy the parallel trends assumption so this assumption can still be reasonable. It is important to examine this variable since test scores are a key metric for school performance and the magnitudes of the estimated coefficients seem sensible. We can compare results to Baron (2018) who found that composite scores declined by 22% of a standard deviation following exposure to Act 10. Instead, this model suggests proficiency in composite tests contemporaneously increased by 0.961 percentage points, but 2 years after treatment the effect is insignificant. Whilst results from the event study design are not consistent with Baron's (2018) findings, results from the triple difference model do suggest worsening test scores.

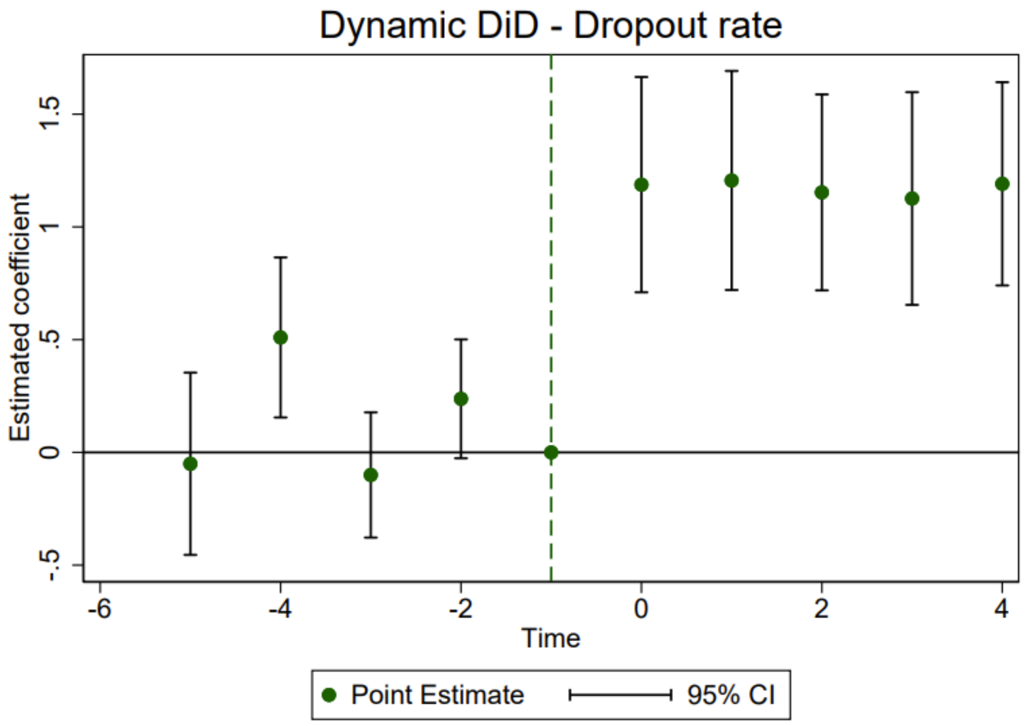

Figure 7: Event study design for dropout rate

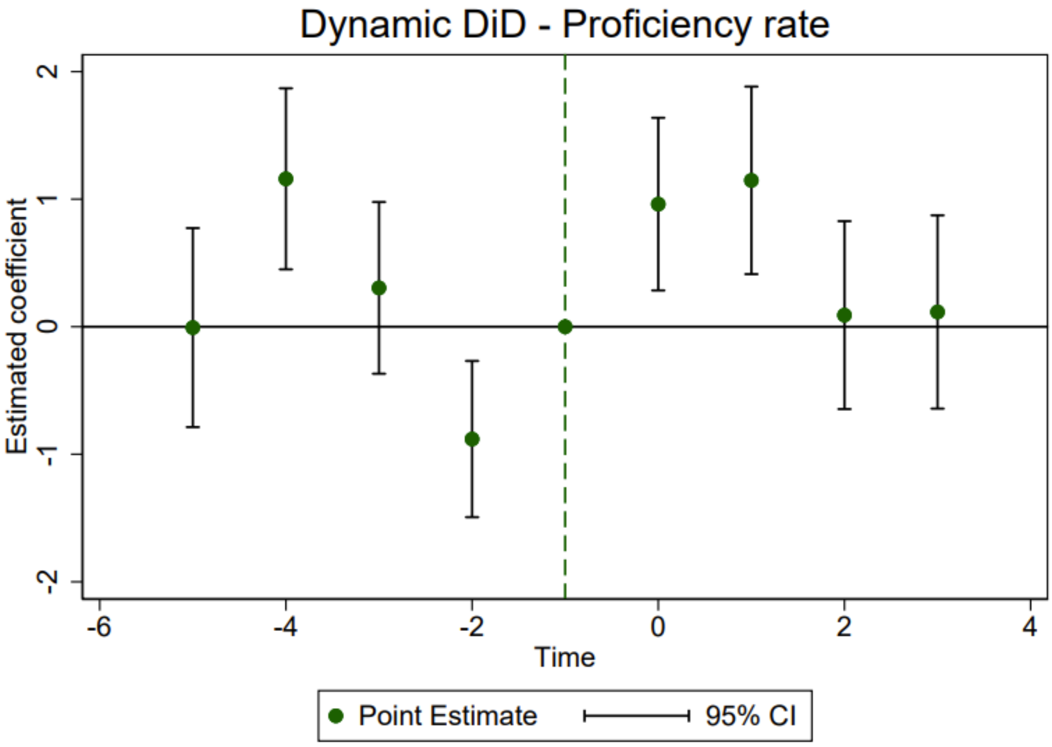

Figure 8: Event study design for proficiency rate

5.2 Triple Difference

Table 3 shows the results of the triple difference estimation. It presents the triple difference estimates and standard errors of the aggregate impact of Act 10. Standard errors are clustered at the school level and models are weighted by school size. Two different specifications are estimated for each outcome variable: a parsimonious model and a model with full controls. Including controls in a difference-in-difference model should increase statistical power. This makes it more likely a significant result is found when one exists.

Table 3: Triple difference model results

VARIABLES	(1) Dropout rate	(2) Dropout rate	(3) Proficiency rate	(4) Proficiency rate
DDD estimate	-3.328	-1.599	-9.409**	-11.91***
	(2.916)	(3.212)	(4.226)	(4.026)
Controls[1]	No	Yes	No	Yes

[1]Note: even numbered models include controls for school characteristics (percent White, Black, Hispanic, Asian) and district level characteristics (local property taxes and total revenue). Standard errors are clustered at school level and all models are weighted by school size. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Compared to the event study design, the dropout rate estimate is now insignificant and negative. However, the event study's coefficient estimates fall within the confidence interval for the triple difference model's estimates.

Coefficient estimates for the proficiency rate are statistically significant for both the parsimonious model and model with full controls. This model suggests that the proficiency rate fell 11.91 percentage points on average for schools exposed to Act 10. This implies that weakening teachers' unions can have a negative impact on test scores, which is consistent with Baron (2018).

5.3 Discussion

A difference-in-difference model was estimated using public schools and charter schools in Wisconsin (table A.6) to give an insight into the observed results. The estimate for the dropout rate is insignificant but finds that the proficiency rate for public schools was negatively affected by Act 10 compared to charter schools at the 10% significance level. This suggests that there may be some within-state spillover effects where better teachers switched from public to charter schools. This is supported by the fact that we see a decline in teacher quality (measured by experience and qualification level) in Wisconsin's public schools following Act 10 (see online appendix). Roth (2019) documents a short-run increase in teacher turnover in public schools after the Act.

We see that dropout rates increased for Wisconsin relative to Illinois, but not between charter and public schools in Wisconsin. This could be because teachers moving to charter schools had trouble adapting to the different school culture which could encourage students in charter schools to drop out on the margin. This increase in charter school dropout rates would counteract the increase in dropout rates seen in public schools, resulting in an insignificant estimate from the triple difference model. The coefficients from the event study design fall within the confidence intervals for the triple difference model which suggests that the magnitudes are plausible. Thus, dropout rates likely increased in Wisconsin's public schools due to Act 10.

We also see that changes in proficiency rates were relatively insignificant for public schools in Wisconsin compared to Illinois, but that proficiency rates fell for public schools compared to charter schools. This is likely because good teachers moved from public to charter schools which improved the relative test scores for charter schools. The triple difference model can amplify the effect of the shock to charter schools, thus finding a large effect. Results from the triple difference model are robust to differences in economic conditions between states, so more confidence should be placed in its findings. Thus, it is likely that test scores declined for Wisconsin following Act 10, which is consistent with Baron (2018).

5.4 Robustness Checks

A concern using aggregate school-level data is that school composition may have changed following Act 10. For instance, if some parents were concerned by the disruption caused by the reforms, they might have removed their children from the affected schools and enrolled them in unaffected schools. If this were the case, estimates may be driven by a change in the composition of students in treated schools rather than the direct effects of Act 10. To test whether the reforms led to significant changes in school composition, an event study design is estimated with the percentage of white students, the percentage of black students, and the school's total enrollment as the outcomes of interest. These results are shown in table A.4 and figures A1-3 in the online appendix. All post-treatment coefficients for these variables are insignificant, apart for the event time 2 and 3 coefficients for the percentage of Black students which are marginally significant at the 5% level and insignificant at the 1% level. Overall, this suggests that there were no significant changes in school composition following Act 10.

All models are estimated using the matched dataset and pre-match dataset to ensure results are robust to matching. Output tables in the online appendix (A.3 and A.5) show results for the event study design and triple difference model estimated using the pre-match dataset. For both models, the magnitudes of results are similar, suggesting that results are robust to the dataset used for estimation.

6. Limitations

The main limitation is one of external validity. The educational effects of antiunion laws likely depend on the specifics of the policy in question and institutional structures. For example, states may try to cut costs in different ways which may affect teachers' working conditions differently. It would be interesting to expand this analysis to a country-wide level to examine on aggregate the effects of weakening teachers' unions. Lack of complete data for educational outcomes in Idaho, Indiana, and Tennessee limited the scope of this study to Wisconsin's Act 10 rather than the antiunion laws implemented in these states.

Furthermore, the outcome variables considered are not a perfect measure of school success. Other variables, such as the out-of-school suspension rate, could be considered to give an even broader view of educational outcomes. Changes to the truancy rate definition limit the scope of analysis for this variable. Test scores can be a problematic measure of success when considered alone. Reliance on test-based school accountability can incentivize teachers to cheat by changing students' answers (Jacob and Levitt, 2003) or to "teach to the test" (Neal, 2012) which may limit true student development.

Further study could examine in more detail the likely mechanisms behind the observed trends in educational outcomes. Whilst raw plots over time can give an indication as to the correlation between the secondary outcome variables and educational outcomes (online appendix), it must be made clear that this does not imply a causal effect.

7. Conclusion

This study examines the impact of weakening teachers' unions on educational outcomes. It does so by exploiting differences in exposure to Act 10, a law that significantly limited the power of teachers' unions in Wisconsin. A unique school-level dataset is constructed containing data on high schools from Wisconsin and Illinois between 2006-2015. A difference-in-difference methodology is used to mitigate problems surrounding simultaneity and omitted variable bias from standard OLS regression and propensity score matching is used to create a more valid control group. The triple difference model used has lower power but is more robust to the parallel trends assumption as it takes into account differences in economic conditions between states. This paper provides evidence that weakening teachers' unions can have a negative impact on the student dropout rate and test scores, supporting the union-voice hypothesis. Differences in model findings can be reconciled by within-state spillover effects where good teachers in Wisconsin left public schools due to Act 10. The negative impact of Act 10 on educational outcomes is consistent with Baron (2018) which increases confidence in the results.

Bibliography

Alexander, K., Entwisle, D. and Bedinger, S. (1994). "When Expectations Work: Race and Socioeconomic Differences in School Performance." *Social Psychology Quarterly*, 57(4): 283-299.

Baron, J. (2018). "The Effect of Teachers' Unions on Student Achievement in the Short Run: Evidence from Wisconsin's Act 10." *Economics of Education Review*, 67: 40-57.

Baron, J. (2021). "Union Reform, Performance Pay, and New Teacher Supply: Evidence from Wisconsin's Act 10." *AEA Papers and Proceedings*, 111: 445-49.

Bau, N. and Das, J. (2020). "Teacher Value Added in a Low-Income Country." *American Economic Journal: Economic Policy*, 12(1): 62-96.

Becker, G. S. (1993). "Human capital: A theoretical and empirical analysis with special reference to education." Chicago: University of Chicago Press.

Bertrand, M., Duflo, M. and Mullainathan, S. (2004). "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119 (1): 249–75.

Biasi, B. (2018). "The Labour Market for Teachers Under Different Pay Schemes." NBER Working Paper No. 24813.

Bradley, S. and Taylor, J. (2001). "The Effect of School Size on Exam Performance in Secondary Schools." *Oxford Bulletin of Economics and Statistics*, 60(3): 291-324.

Britton, J. and Propper, C. (2016). "Teachers pay and school productivity: Exploiting wage regulation." *Journal of Public Economics*, 133: 75-89.

Bureau of Economic Analysis. (2022). "Regional Data: State GDP." Available at: https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1&acrdn=1 [Last accessed 25/02/2022].

Card, D., Lemieux, T. and Riddell, C. (2003). "Unionization and Wage Inequality: A Comparative Study of the U.S, the U.K., and Canada." NBER Working Paper No. 9473.

Carl, B. and Caldas, S. (1997). "The American School Dilemma: Race and Scholastic Performance." *The Sociological Quarterly*, 38(3): 423-429.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W, and Yagan, D. (2011). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics*, 126 (4): 1593–1660.

Clarke, D. and Schythe, K. T. (2020). "Implementing the panel event study." *Institute of Labor Economics*, Discussion Paper Series: 13524.

Cutler, D.M. and Lleras-Muney, A. (2006). "Education and Health: Evaluating Theories and Evidence." NBER Working Paper No. 12352.

DiSalvo, D. (2015). "Government against Itself: Public Union Power and Its Consequences." Oxford: Oxford University Press.

Doucouliagos, C. and Laroche, P. (2003). "What Do Unions Do to Productivity? A Meta-Analysis." *Industrial Relations,* 42(4): 650-691

Easterbrook, J., Kuppens, T., and Manstead, R. (2016). "The education effect: higher educational qualifications are robustly associated with beneficial personal and socio-political outcomes." *Social Indicators Research,* 126(3): 1261-1298.

Frandsen, B. (2016). “The Effects of Collective Bargaining Rights on Public Employee Compensation: Evidence from Teachers, Fire Fighters, and Police.” *Industrial and Labor Relations Review*, 69(1): 84-112.

Freeman, R. (1980). “The Exit-Voice Tradeoff in the Labor Market: Unionism, Job Tenure, Quits, and Separations.” *Quarterly Journal of Economics*, 94(4): 643-673.

Freeman, R. and Han, E. (2012). “The war against public sector collective bargaining in the US.” *Journal of Industrial Relations*, 54(3): 386–408.

Freeman, R. and Medoff, J. (1985). “What do unions do?” *ILR Review*, 38(2): 244–263.

Goldhaber, D. and Brewer, D. (2000). “Does teacher certification matter? High school teacher certification status and student achievement.” *Educational Evaluation and Policy Analysis*, 22(2): 129-146.

Greenwald, B. and Stiglitz, J. (1988). “Pareto inefficiency of market economies: search and efficiency wage models.” NBER Working Paper No. 2651.

Grossman, M. (2006). "Education and Nonmarket Outcomes." Handbook of the Economics of Education, Elsevier.

Gunderson, M. (2005). “Two Faces of Union Voice in the Public Sector.” *Journal of Labor Research*, 26(3): 393-413.

Hammer, T.H., Avgar, A. (2005). “The impact of unions on job satisfaction, organizational commitment, and turnover.” *Journal of Labor Research*, 26(2): 241–266.

Helliwell, J. and Putnam, R. (2007). “Education and Social Capital.” *Eastern Economic Journal*, 33: 1–19.

Hendricks, M. (2014). “Does it pay to pay teachers more? Evidence from Texas.” *Journal of Public Economics*, 109: 50-63.

Hensvik, L. (2012). “Competition, Wages and Teacher Sorting: Lessons Learned from a Voucher Reform.” *The Economic Journal*, 112: 799-824.

Imai, K., and Kim, I. (2021). “On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data.” *Political Analysis*, 29(3), 405-415.

Jacob, B. A. and Levitt, S. D. (2003). “Rotten apples: An investigation of the prevalence and predictors of teacher cheating.” *The Quarterly Journal of Economics*, 118(3): 843–877.

Kaufman, D. (2021). “Scott Walker’s Wisconsin Paved the Way for Donald Trump’s America.” *The New York Times*. Available at: https://www.nytimes.com/2021/07/06/opinion/wisconsin-act-10-trump.html [Last Accessed 08/02/2022].

Lehr, C. A., Johnson, D. R., Bremer, C. D., Cosio, A., and Thompson, M. (2004). “Essential tools: Increasing rates of school completion: Moving from policy and research to practice.” Minneapolis, MN: University of Minnesota, Institute on Community Integration, National Center on Secondary Education and Transition.

Lewit, E. (1992). “Dropout rates for high-school students.” *The future of children*, 2(1): 127-130.

Lochner, L. and Moretti, E. (2004). "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review*, 94 (1): 155-189.

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

Mbiti, I., Romero, M. and Schipper, Y. (2019). “Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania.” NBER Working Paper No. 25903.

Mincer, J. (1974). “Schooling, experience and earnings.” New York: Columbia University Press.

Meng, R. (1990) “The relationship between unions and job satisfaction.” *Applied Economics*, 22(12): 1635-1648.

Murphy, M. and Zirkel, S. (2015). “Race and Belonging in School: How Anticipated and Experienced Belonging Affect Choice, Persistence, and Performance.” *Teacher College Record*, 117(12): 1-40.

Neal, D. (2012). “Providing incentives for educators.” Chapter 4 in Handbook of Economics of Education, E. Hanushek, S. Machin, and L. Woessmann, eds. Elsevier.

Olden, A. and Moen, J. (2020). “The Triple Difference Estimator.” *NHH Dept. of Business and Management Science*, Discussion Paper No. 2020/1.

Oreopoulos, P. and Salvanes, K. (2011). "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives*, 25(1): 159-84.

Oswald, A. J. (1985). “The Economic Theory of Trade Unions: An Introductory Survey.” *The Scandinavian Journal of Economics*, 87(2), 160–193.

Papay, J. and Kraft, M. (2015). “Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement.” *Journal of Public Economics*, 130: 105-119.

U.S. Bureau of the Census. (1991). “Statistical abstract of the United States.” 111th ed. Washington, DC: U.S. Government Printing Office.

Rauch, J. (1993). “Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities.” *Journal of Urban Economics*, 34(3): 380-400.

Rosenbaum, P. and Rubin, D. (1983). “The central role of the propensity score in observational studies for causal effects.” *Biometrika*, 70(1): 41–55.

Rosenfeld, J. and Denice, P. (2019). “What do government unions do? Public sector unions and nonunion wages, 1977–2015.” *Social Science Research*, 78: 41-56.

Roth, J. (2019). Union Reform and Teacher Turnover: Evidence from Wisconsin’s Act 10. Working Paper.

Rumberger, R. W. and Palardy, G. J. (2005). “Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance.” *American Educational Research Journal*, 42(1): 3–42.

Schultz, T. (1961). “Investment in human capital.” *American Economic Review*, 51: 1–17.

Stiefel, L., Berne, R., Latarola, P. and Fruchter, N. (2000). “High School Size: Effects on Budgets and Performance in New York City.” *Educational Evaluation and Policy Analysis*, 22(1): 27-39.

Stuart, E.A., Huskamp, H.A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. and Barry, C. (2014). “Using propensity scores in difference-in-differences models to estimate the effects of a policy change.” *Health Services and Outcomes Research Methodology*, 14: 166–182.

Thurlow, M. L., Sinclair, M. F., and Johnson, D. R. (2002). "Students with disabilities who drop out of school: Implications for policy and practice." Institute on Community Integration, National Center on Secondary Education and Transition, University of Minnesota.

Venteicher, J. (2005). "How Much Does Funding Matter? An Analysis of Elementary and Secondary School Performance in Missouri, 1990-2004." *Journal of Educational Research and Policy Studies,* 5(2): 39-65.

Villa, J. M. (2016). "diff: Simplifying the estimation of difference-in-differences treatment effects." *Stata Journal*, 16: 52-71.

Wisconsin Policy Forum. (2021). "Fewer Education Grads Pose Threat to Teacher Workforce." Available at: https://wispolicyforum.org/research/fewer-education-grads-pose-threat-to-teacher-workforce/ [Last Accessed: 08/02/2022].

Wooden, H. and Hawke, A. (2000). "Unions and the Employment Growth: Panel Data Evidence." *Industrial Relations*, 39(1): 88-107.