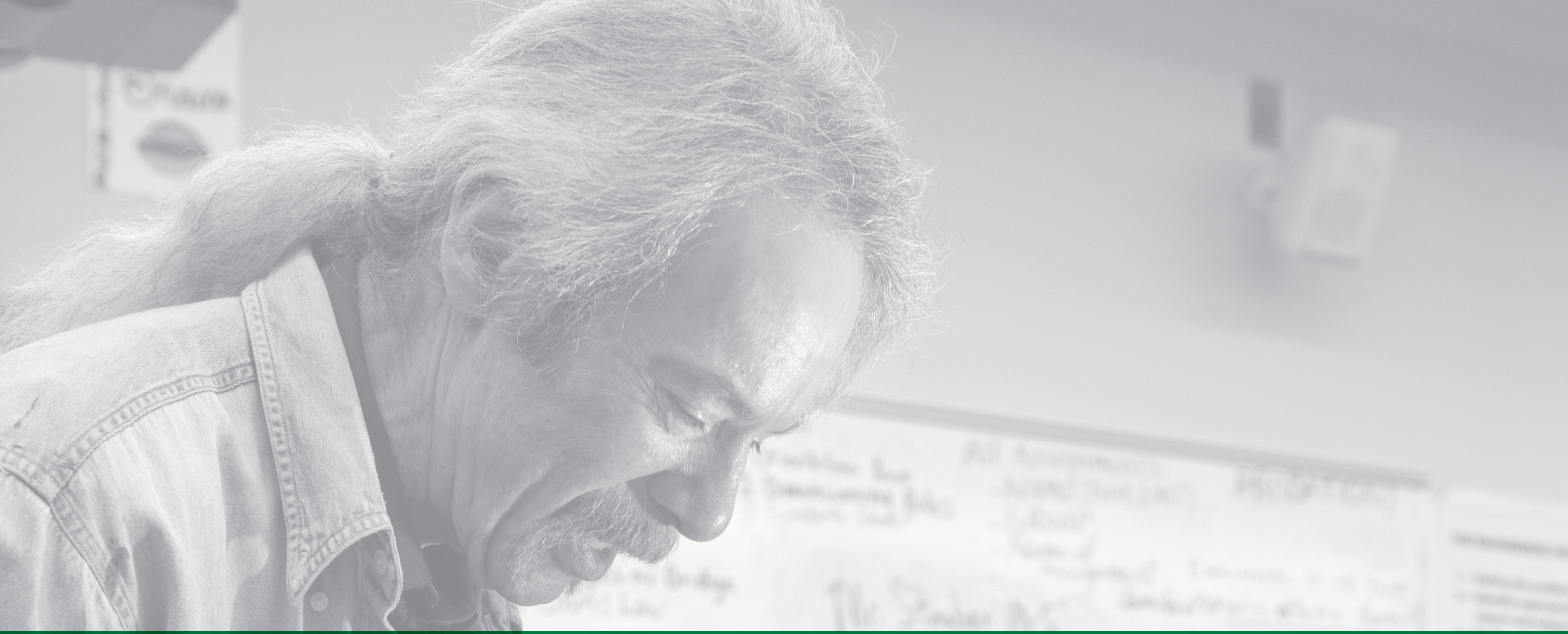


EPIC
Education Policy
Innovation Collaborative
RESEARCH WITH CONSEQUENCE

WORKING PAPER

The Impact of Teacher Labor Market Reforms on Student Achievement: Evidence from Michigan

Kaitlin P. Anderson, *Michigan State University*
Joshua M. *Cowen, Michigan State University*
Katharine O. Strunk, *Michigan State University*

June 2019

The Impact of Teacher Labor Market Reforms on Student Achievement: Evidence from Michigan

Kaitlin P. Anderson
Michigan State University

Joshua M. Cowen
Michigan State University

Katharine O. Strunk
Michigan State University

Abstract

Many states have recently enacted substantial reforms to teacher policies. In Michigan, the state legislature implemented teacher evaluation requirements, reduced tenure protections, and restricted teachers' unions' abilities to collectively bargain. Some teacher advocates view such reform as a "war on teachers," but proponents argue these policies may enable personnel decisions that positively impact student performance. Evidence to discern between these perspectives remains limited. In this study, we use detailed administrative data from all Michigan traditional public schools from 2005-06 to 2012-13. We estimate event study models for the first two cohorts exposed to these reforms, exploiting the plausibly exogenous timing of collective bargaining agreement expirations. Across a variety of samples and specification checks, we find very little evidence of negative achievement impacts, with generally null impacts on the first cohort exposed to reform, and generally positive impacts on the second cohort exposed to reform.

Keywords: teacher quality; teacher evaluation; collective bargaining; tenure protections; teacher labor markets; education policy; student achievement

JEL classifications: I21, I24, I28

Support for this project is provided by the Laura and John Arnold Foundation and an anonymous foundation. This research uses data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information and opinions solely represent the analysis, information and opinions of the author(s) and are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof. We thank Eric Brunner for his helpful comments on earlier versions of this work. All errors are our own.

1. Introduction

In recent years, states have enacted a wave of broad teacher labor market reforms that include high-stakes teacher evaluation based in part on student test scores, the linking of tenure to teacher effectiveness, and the weakening of teachers' collective bargaining rights. Supporters of such reforms argue that these policies empower school and district leaders to make personnel decisions that positively impact students. Teacher advocates, on the other hand, have viewed these changes as part of a "war on teachers," and on teachers' unions in particular, arguing that such policies damage the teaching profession, reduce the supply of effective teachers, and ultimately impede student outcomes. These debates have become more pronounced with the U.S. Supreme Court's ruling in *Janus v. AFSCME* that it is unconstitutional for unions to collect agency fees from employees as a condition of employment (*Janus v. AFSCME, 2018*). That ruling, which substantially limited unions' ability to raise revenues and—potentially—maintain member services, followed a series of teacher strikes and strike threats, as teachers protested stagnant salaries and worsening working conditions.

Despite the intensity of both support for and opposition to these teacher-related policy changes, there is little empirical evidence concerning the impact of such reforms. In this paper, we focus on one key outcome—student achievement—to provide such evidence. We focus on a set of reforms in Michigan that in 2011 and 2012 increased evaluation requirements for public school teachers, reduced tenure protections, restricted teachers' unions' abilities to collectively bargain and—in a foreshadowing of the *Janus* ruling—removed teachers' unions' abilities to collect agency fees. We use flexible event-study models to estimate the impact of exposure to these reforms over time. This strategy relies on the fact that teachers were not subject to the main slate of reforms until their districts' pre-reform collective bargaining agreement (CBA) expired. We use this differential timing of reform exposure to construct comparison groups that differ for each exposure cohort. We find little evidence that these reforms impacted student achievement for the first cohort of students exposed to these reforms, but find evidence that the second cohort of students benefited in both ELA and math, with economically disadvantaged benefiting in particular.

The remainder of the paper proceeds as follows: in Section 2, we describe the set of teacher policy reforms in Michigan, and in Section 3, we review the available evidence on the potential for these types of reforms to affect student achievement. Section 4 describes our data, empirical framework, and samples. In Section 5, we present our results, and finally, in Section 6, we discuss the major implications of this work moving forward.

2. Background: Teacher Labor Market Reforms in Michigan

In 2011 and 2012, Michigan implemented a series of reforms that substantially reduced teachers' tenure protections and the rights of teachers' unions to collectively bargain over certain conditions of employment. The first wave of reforms came in July 2011, when the Michigan legislature passed Public Acts 100, 101, 102, and 103 (State of Michigan, 2011). A major focus of reform was Michigan's existing teacher evaluation system. Combined, Public Acts 100 through 103 replaced the existing system with a high stakes performance-based teacher evaluation system and tied promotion and layoff decisions to evaluation measures. Public Act 102 established a teacher evaluation system in which teacher evaluations were required to include student achievement as a "significant" determinant of educator performance ratings and teachers with three "ineffective" ratings were to be dismissed.[1]

[1] Later, in November 2015, districts were allowed wide discretion in the implementation of this policy over time, although student achievement remained an important feature.

Public Acts 100 through 102 also tied evaluation outcomes and teacher effectiveness to staffing decisions around layoffs and tenure. Public Act 102 required districts to base layoff decisions on the new evaluation ratings and prohibited the use of seniority as the primary determinant in layoff decisions. Under the new policy, which essentially prohibited Last-In First-Out layoff procedures, seniority can only be used to break ties between similarly rated teachers. Public Act 101 increased the pre-tenure probationary period from four to five years and required that tenure decisions be tied to teacher effectiveness by mandating three consecutive years of effective ratings or higher before receiving tenure (State of Michigan, 2011). Public Act 100 allowed for removal of tenure protections for reasons "not arbitrary or capricious." Public Act 103 restricted unions' collective bargaining rights. In particular, PA 103 prohibited new CBAs from governing issues such as teacher evaluation, transfer, reassignment, performance-based compensation, the length of the school year, and discipline (State of Michigan, 2011). The 2011 reforms were followed by Public Act 349 in December 2012, which, foreshadowing the *Janus* decision six years later, made it unlawful for Michigan school districts to require teachers to pay agency fees (i.e., union dues) as a condition of employment. Thus, Public Act 349 made Michigan a Right-to-Work state (State of Michigan, 2012).

The timing of these reforms is important for our purposes here: Public Acts 100 and 101 theoretically applied to new teachers immediately (although the requirement to base tenure restrictions on teaching effectiveness may have been limited by the lack of rigorous evaluation until after the process was clarified by Public Act 173 of 2015).[2] However, Public Acts 102 and

[2] Although all districts began evaluating teachers by the fall of 2011, there was substantial variation in implementation across districts in terms of timing and rigor. Public Act 173 of 2015 clarified some of the requirements surrounding teacher evaluations to ensure the process was more "rigorous, transparent, and fair," and it was not until 2016-17 that school districts and public school academies (i.e. charter schools) were required to train all teachers, administrators, evaluators, and observers on the observation tools used to evaluate teachers (State of Michigan, 2015). As a result, it is possible that the impact of the initial 2011 reforms may not have been felt immediately, at least with respect to how teacher performance was evaluated.

103 did not take effect until the previously bargained CBAs expired, which occurred at different times in different districts. As a result, teachers protected by existing CBAs from achievement-based dismissal were not impacted by evaluation reforms, and collective bargaining was not constrained until current CBAs expired. Public Act 349 went into effect in March 2013.[3] All told, the full battery of these teacher-related policy changes "radically altered the landscape of bargaining for public school employers and the unions representing their teachers" (Michigan Association of School Boards, n.d.).

For a timeline of political events, and how they correspond to the timing of reform exposure across districts, see Figure 1.

3. Literature Review

As changes similar to Michigan's have been implemented across the country, a growing body of scholarship is emerging that assesses the impacts of such policies on student achievement. This literature, while relatively nascent, suggests that teacher labor market reforms of the sort implemented in Michigan may lead to improvements in student achievement. For instance, studies that examine the effects of teachers' participation in high-stakes rigorous teacher evaluation systems in Cincinnati, Washington, D.C., and Chicago find that participation in teacher evaluation systems – combined with intensive supports (as in Cincinnati Public Schools and Chicago Public Schools) or supports alongside both positive incentives for effectiveness and negative consequences for ineffectiveness (as in Washington, D.C.) – lead to improvements in the achievement of participating teachers' students (Dee & Wyckoff, 2015; Steinberg & Sartain, 2015; Taylor & Tyler, 2012).

[3] As with PA 102 and 103, PA 349 did not apply until existing pre-reform contracts expired. In most cases, these pre-PA 349 contracts were signed after PA 100-103 implying two sets of pre and post-reform contracts. In the identification strategy employed in this study, however, we focus only on the expiration of the contracts signed prior to July 2011, when the initial wave of reform occurred.

While some studies have looked at the effects of removing or limiting tenure protections on teacher exit (Strunk, Barrett, & Lincove, 2017; Brunner, Cowen, Strunk, & Drake, forthcoming) there have been fewer studies linking the removal of tenure protections directly to student outcomes. In the only study of which we know on the latter topic, Carruthers, Figlio, and Sass (2018) estimate the short-run student achievement impacts of the full removal of tenure protections in Florida in 2011. They find little reform effect on student achievement overall, but some evidence of gains for the quintile of students who were lowest performing in math and reading pre-reform.

Additionally, recent work that assesses the impact of seniority-based layoffs on student achievement provides evidence that shifting layoff decisions from a reliance on seniority to other merit-based considerations may improve student outcomes (Kraft, 2015; Strunk, Goldhaber, Knight, & Brown, 2018).

Reforms that weaken or remove collective bargaining protections might improve student achievement if those teacher protections were previously hindering student outcomes. However, because there is no extant work that specifically examines whether the removal or weakening of collective bargaining rights is associated with changes in student achievement, such a possibility can be only inferred through evidence of the impact of the presence of unions, bargaining, or contract strength on student outcomes. That evidence, which has been slowly accumulating over the past two decades, generally suggests that stronger unions lead to lower student achievement and worse long-term outcomes (e.g., Hoxby, 1996; Lovenheim & Willen, 2018; see Cowen & Strunk, 2015 for a review). Studies focusing on the relationship between the restrictiveness of teachers' union contracts and student achievement similarly suggest null to negative impacts of CBA strength on student outcomes (Marianno & Strunk, 2018; Moe, 2009; Strunk, 2011; Strunk & McEachin, 2011).

Such studies can inform expectations about the impact of Michigan's reforms on student achievement, but they do not directly consider such reforms themselves. New research on the impact of Wisconsin's Act 10, which greatly limited the bargaining power of teachers' unions in the state, provides the only direct evidence of such reforms on achievement outcomes prior to our present study. As in Michigan, the reforms primarily affected districts after the expiration of their existing CBAs. Perhaps due to different data granularity, grade levels of focus, and analytic approaches, two papers on Wisconsin's Act 10 find different results: an improvement in achievement, in grades where teachers retired (Roth, 2019), or a decrease in achievement, primarily in low-performing schools (Baron, 2018). Roth (2019) estimates school-by-grade value-added in grades 3-5, focusing on these grades due to the availability of yearly testing and the prevalence of self-contained classrooms. He finds an increase in teacher turnover after the reform, primarily driven by retirements of older teachers, who were facing strong incentives to retire prior to the expiration of the CBAs in their districts. Roth also finds that student achievement improved in the school-grade combinations in which teachers retired. However, Baron (2018), using school-level averages and estimating impacts on tenth grade math achievement, finds that test scores decreased by about 20% of a standard deviation overall after the reform, largely driven by declines in low-performing schools. He argues that his findings may be due primarily to a large increase in teacher turnover. Due to these varied findings, these two Wisconsin studies—while providing important steps forward in the research on reform impacts—provide no clear indication of what to expect in Michigan.

4. Data and Empirical Framework

4.1 Data

We use detailed administrative records for traditional public school students provided by the Michigan Department of Education (MDE) and the Center for Educational Performance and Information (CEPI) from 2005-2006 to 2012-13.[4] These administrative records include student demographic information and student achievement on the Michigan Educational Assessment Program (MEAP) from 2005-06 to 2012-13.[5] Students in grades 3-8 took tests in both mathematics and reading. These test scores are standardized within grade level, subject, and academic year.[6]

4.2 Empirical Framework

As noted above, Public Acts 101-103 prohibited new CBAs (those bargained after July of 2011) from governing policies related to evaluation, teacher transfer and reassignment, performance-based compensation, classroom observations, the length of the school year, and teacher discipline. However, because CBAs are locally negotiated, expiration dates vary across districts, and districts were fully subject to Michigan's 2011 reforms at different times, depending on when the pre-reform CBA negotiated by that school district expired.[7] Approximately 40% of districts had pre-reform CBAs that expired in 2011, making those districts immediately susceptible to the reforms. Another 29% of districts' pre-reform CBAs expired in 2012, 11% expired in 2013, and 2% expired in either 2014 or 2015. Finally, 10% of

[4] Data were available through 2016-17, but we restrict the analysis to 2012-13 in order to only include years with a sufficiently large comparison group of districts yet to be exposed to reform. Only 8 districts remained unexposed in 2013-14, so we restrict the analytic frame to 2012-13 and earlier.
[5] MEAP testing occurred in the fall, so a test taken in fourth grade, for example, covered third grade standards.
[6] To create standardized test scores, we include the entire population of Michigan public schools, including charter schools, even though our analytic sample focuses on traditional public schools.
[7] The state began requiring districts to begin teacher evaluations in the fall of 2011, however the consequences outlined in the reform legislation did not take effect until districts' CBAs expired.

districts had CBAs that expired in 2010, and 9% had previously expired contracts with no new contracts in place.

In essence, our initial strategy relies on the notion that districts whose pre-reform (i.e. pre-2011) contracts were signed prior to December 2009—in some cases a number of years prior—would have been unable to anticipate the precise timing of legislation passage in July 2011. Thus, the expiration dates of those last pre-reform contracts were idiosyncratic to individual districts, with the possible exception of districts whose expiration dates occurred near enough to the 2011 reform. After consulting with the staff at Michigan Department of Education, who suggested districts with contracts that expired later than December 2009 may have acted based on anticipated changes in the law, our main specifications include the 2009 and 2010 expirations as occurring in 2011, such that 59% of districts are coded as treated in 2011.[8] Thus, the 2009 and 2010 expirations can be considered "intent-to-treat" districts tied to the 2011 expiration. In addition, relative to simply estimating the post-2011 change, a final advantage of this approach is that the timing of treatment exposure varies across districts, which reduces the potential for other contemporaneous factors, such as the Great Recession, to confound our estimates.

Given that the large majority of districts were either in the 2011 exposure cohort (40%) or the 2012 exposure cohort (29%), we focus our event study specifications on these two groups of districts, forming a comparison group from the later cohorts not yet exposed. For example, reform effects on the 2011 exposure cohort are estimated using the 2012 and later cohorts as the comparison group, and the reform effects on the 2012 exposure cohort are

[8] In alternative specifications, we estimate the effects on the 2009 and 2010 cohorts separately and also estimate the effects on the 2011 cohort excluding the 2009 and 2010 cohorts.

estimated using the 2013 and later cohorts as the comparison group. In each of these cases, comparison group observations that occur after that cohort is exposed are dropped from the analysis, using only their pre-treatment observations to construct the comparison group observations. For example, in the 2011 exposure cohort analysis, the 2012 exposure cohort is observed as the counterfactual in 2011-12 and earlier (before it is treated), but not in 2012-13, when it begins being treated. In 2012-13, the comparison group only includes the 2013 exposure cohorts and later. In addition, since the vast majority of districts had already been exposed by 2013-14, and in order to retain a clean counterfactual of untreated comparison districts, we restrict the analytic frame to 2012-13 and earlier.

Our event-study strategy exploits this plausibly exogenous timing of CBA expiration dates, taking the form:

$$(1) \qquad Y_{it} = \beta_0 + \sum_{r=-3}^{2} Exposed_{it} + \boldsymbol{X_{it}\gamma} + \boldsymbol{\lambda_t} + \boldsymbol{\theta_i} + \varepsilon_{it}$$

Where Y_{it} is a potential student achievement outcome (either math or ELA) for student *i* in time *t*. The term $\sum_{r=-3}^{2} Exposed_{it}$ indicates a series of indicators for the number of years since exposure to the reform (defined by the timing of the expiration of the pre-reform contract), where the omitted category is the year just prior to exposure (for example, we omit 2010 for the 2011 exposure cohort's models). Specifically, we assume that a student is first "exposed" in the first year in which he or she was in a district in which the CBA had previously expired. In other words, once a student is "exposed" to a new contract, he or she remains always exposed. The coefficients on the pre-reform exposure year indicators show whether pre-reform trends were different in the treatment and comparison cohorts. For a causal interpretation of our results, these pre-reform interactions should be small and statistically insignificant. The coefficients on the post-reform exposure year indicators provide evidence about the impact of reform in each year following reform exposure, although this is limited to only one outcome year for the 2012 exposure cohort.

We control for unobservable characteristics of students that are time-invariant using student fixed effects, $\boldsymbol{\theta_i}$, and $\boldsymbol{X_{it}}$ is a vector of student time-varying characteristics including indicators for economically disadvantaged status,[9] English proficiency status, special education status, and a set of indicators for each grade level. For students who remain in one district over our panel (the vast majority, 84% of students), we allow the student fixed-effect to subsume a district fixed effect as well.[10] $\boldsymbol{\lambda_t}$ are academic year fixed effects. Finally, ε_{it} is a random disturbance term. We cluster our standard errors at the district level to allow for within-district autocorrelation of the disturbance term.

While these models capture the policy impact on student achievement overall, we are also interested in whether there are heterogeneous impacts across groups of students. Thus, we estimate similar models separately for economically-disadvantaged students and underrepresented minorities (non-White, non-Asian).

4.3 Sample and descriptive statistics

Our preferred sample excludes the Detroit public schools because the Detroit school system is an outlier in Michigan in many ways. The decline of the auto industry had a particularly devastating impact on the Detroit area, and in 2009, Detroit public schools faced a $400 million deficit, forcing the school system into emergency management. The emergency management is especially germane to our identification approaches here, because under Public

[9] FRL-eligibility for each student was indicated for 2005-06 through 2011-12. Starting in 2012-13, Michigan reported student economic disadvantage which includes FRL-eligibility as well as other indicators of disadvantage such as homelessness, migrant status, foster status, and receipt of TANF or SNAP benefits.
[10] Our results below are robust to excluding and including movers in the sample and to the inclusion of a district fixed effect.

Act 4 of 2011, and later under Public Act 436 of 2012, the emergency manager had the power to "reject, modify, or terminate one or more terms and conditions of an existing collectively bargained agreement" (State of Michigan, 2011, 2012).

Table 1 presents summary statistics for our samples. We describe our preferred samples, excluding Detroit, in Panel A, and our samples including Detroit in Panel B. Across the four columns, representing the math and ELA samples for estimating the impact on the 2011 and 2012 exposure cohorts, we can see that the 2011 Exposure Cohort samples are larger, as expected, because the 2011 cohort is dropped from the 2012 sample. Student-year observations in the 2012 Exposure Cohort sample tend to be in slightly higher grades, on average. Also, the 2012 samples include a larger share of special needs students. There are some differences when including Detroit, all in expected ways. In particular, the sample including Detroit is larger, lower performing, has more non-White non-Asian students, and more economically disadvantaged students.

While we argue that the expiration of the pre-reform CBA should have been determined well in advance of the new legislation, and that as a result, the timing of these CBA expirations was essentially random, we also test for differences across cohorts by comparing the characteristics of the districts in each exposure cohort. Included in Table 2 are a variety of demographic characteristics, student achievement, teacher characteristics, and community characteristics, in 2010. We include the seven districts exposed in 2014 and the single district exposed in 2015 with the 56 districts exposed in 2013. This group of 64 districts exposed in 2013 and later contains fewer districts and is relatively distinct from the other exposure cohorts. In particular, this cohort includes larger, more suburban, districts with greater numbers of teachers, particularly when compared to the 2012 exposure cohort. In addition, this group of districts exposed in 2013 and later is slightly higher performing than the other exposure cohorts.

We also include in Table 2 measures of CBA restrictiveness based on the last pre-reform contract.[11] Specifically, we include measures of restrictiveness overall, as well as five of the key areas that were restricted by PA 103. These measures are generated using a Partial Independence Item Response (PIIR) model developed by Reardon and Raudenbush (2006) and later refined by Strunk and Reardon (2010) to measure latent levels of contract restrictiveness expressed in CBA content.[12] These measures have been validated in other work (e.g., Goldhaber, Lavery, Theobald, D'Entremont, & Fang, 2013; Marianno et al., 2017; Strunk, 2011; Strunk & Grissom, 2010; Strunk & McEachin, 2011). For more information on the methods used to create the restrictiveness measures, readers can refer to Strunk and Reardon (2010) and Marianno and Strunk (2018). For a detailed discussion of the generation of the Michigan CBA dataset, see Marianno et al. (2017). Table 2 indicates that the 2013 and later exposure cohorts had less restrictive CBAs than average, particularly when it comes to economic incentives and compensation, and particularly when compared to the 2012 exposure cohort.

5. Results

5.1 Estimation of policy impact on student achievement

We present the results from our main event study specifications, excluding Detroit, in Table 3 (for the impact on the 2011 exposure cohort) and Table 4 (for the impact on the 2012

[11] We coded CBAs from all unionized districts. 21 (4%) districts were not included because they were not unionized.

[12] The PIIR method models the existence of individual regulations as a function of a contract-specific latent level of restrictiveness, thereby creating a measure of the underlying degree to which CBAs protect teachers by constraining administrators' abilities to set district policy, relative to other CBAs. Rather than simply selecting a set of specific items to create an index for union strength, the PIIR measure is based on a one-dimensional item response theory (IRT) Rasch model which is often used to construct measures of underlying cognitive skills of test takers by predicting the probability that a test taker correctly answers a question of a particular difficulty level (e,g., Hambelton, Swaminathan, & Rogers, 1991; Rasch, 1980; Wright & Masters, 1982).

exposure cohort). The sample in Table 3 includes within the 2011 cohort (the treatment cohort) the districts whose pre-reform CBAs expired in 2009 and 2010. In this sample, the 2011 exposure cohort forms the treatment group, and the 2012 and later cohorts form the comparison group. Recall that all districts are treated, just at different times. Therefore, to obtain a clear counterfactual, we drop all comparison group observations in the year of that cohort's exposure and later.

In Table 3, the coefficients on Treat X 2011 and Treat X 2012 indicate the impact of the reform, in 2011 and 2012 respectively, relative to the reference year, 2010. We see that the coefficients are generally insignificant and close to zero, suggesting that there was no effect of the reform on student achievement overall. We do find a small and marginally significant result for ELA performance in 2011 for economically disadvantaged students, suggesting that there may have been a small negative impact of the reforms on economically disadvantaged students' ELA performance in the year of the reform. The full sample results in the first two columns are also shown in Figures 2 and 3.

The 2012 cohort results are shown in Table 4. In this sample, the 2012 exposure cohort forms the treatment group, and the 2013 and later cohorts form the comparison group. Across all models in Table 4, we estimate increases in math test scores for the cohorts exposed in 2012, suggesting benefits of these reforms overall, as well as for both sub-groups of relatively disadvantaged students. Similarly, we estimate increases in ELA test scores, although the increase for students overall was only marginally significant. Comparing the magnitude of these estimated effects across the different samples, the economically disadvantaged students and non-white non-Asian students are experiencing larger achievement gains post-reform than the full student population, indicating potential for reducing achievement gaps. The results in the first two columns of Table 4, for the full sample, are also shown in Figures 4 and 5.

For a causal interpretation of our results, we must assume that the treatment and comparison groups had parallel pre-reform trends in baseline outcomes. In both Table 3 (2011 treatment cohort) and Table 4 (2012 treatment cohort), the interactions between treatment status and years before 2011 (2012) indicate whether there are concerns about non-parallel pre-trends, which could bias our results. In Table 3, for the samples including all students and economically disadvantaged students, there are no concerns about parallel pre-trends. However, there is an indication that, among the sample of non-White non-Asian students, the treatment group was trending downward, prior to the reform. Thus, for the 2011 cohort, we cannot conclude that the estimated null effect for non-White non-Asian students is a purely causal zero effect, and we cannot say there was no impact on this group, even though we also have no evidence that there was a non-zero impact. As in Table 3 and Appendix Table A, we also have concerns about non-parallel pre-trends for the non-White non-Asian sample shown in Table 4, but this is only marginally significant. Given that this group was trending upward, prior to reform, it is possible that the positive effects estimated for this group are not causal. We graph the coefficients on the Treat X Year variables from the full samples in Figures 2 and 3. These figures reconfirm that in the full sample there is evidence of parallel pre-trends supporting a causal interpretation at least overall.

5.2 Specification checks for 2009 and 2010 exposure cohorts

As discussed above based on conversations with MDE officials, it is possible that districts with contracts that expired later than December 2009 may have acted based on anticipated changes in the law, so our main specifications (as in Table 3 and Figures 2-3) include the 2009 and 2010 expirations as occurring in 2011. These generally null results, in Table 3 and in Figures 2-3, are largely robust to whether the 2009 and 2010 cohorts are included or dropped from the 2011 cohort, as if they had previously been treated. The results of models excluding the 2009 and 2010 cohorts from the 2011 cohort are shown in Appendix Table A and Appendix Figures A and B. There is even more concern about non-parallel trends within the non-White non-Asian sample in the models that exclude the 2009 and 2010 cohorts, preventing a causal interpretation of the results in those models, but across all models, we estimate no change in test scores after exposure to reform.

5.3 Specification check for reference year

We test the sensitivity of our results to using the year prior to reform exposure as the reference year. In our main specifications, we use 2010 as the reference year when estimating effects on the 2011 cohort and use 2011 as the reference year when estimating effects on the 2012 cohort. In Appendix Table B we show the results using reference years one year prior, and the results are similar to our main results: generally null for the 2011 cohort and generally positive for the 2012 cohort.

5.4 Robustness to sample restrictions

A key sample restriction in our case is the exclusion of Detroit from the analysis. As discussed previously, Detroit is an outlier in many ways, most notably because fiscal distress led to the installment of an emergency manager who had the power to reject or modify CBAs. In other words, there were drastic changes in the extent to which Detroit was managed by previously negotiated CBAs that might confound our estimated impact of the set of reforms affecting the entire state. We test whether our results are robust to the inclusion of Detroit and present the results in Appendix Tables C-E. The results generally support what we presented in

Tables 3-4: generally null impacts for the 2011 cohort and generally positive impacts for the 2012 cohort.

6. Discussion and Conclusion

In 2011 and 2012, Michigan experienced a suite of teacher labor market reforms that implemented higher stakes evaluation systems with potential consequences for teachers, altered the scope of collective bargaining, and diminished the unions' ability to raise funds through required union dues. While there is evidence and theory to suggest that these reforms would reduce some of the employment benefits, and as a result, some of the incentive to be a teacher in the state, perhaps leading to teacher churn and discontent, there is also reason to think that the changes in teacher behavior caused by these reforms might serve to improve student achievement, at least in the long run.

What we find, in this case, across a variety of specifications, sample restrictions, and sensitivity checks, is very little evidence to suggest that Michigan's teacher labor market reforms harmed student achievement, with reason to think that the impact may in fact be positive, at least for the second cohort of exposed students. In particular, we find that students who were immediately exposed to the reform through the expiration of school district CBAs in 2011 experienced minimal changes in math or ELA test scores in the two years following reform exposure. There is some evidence of a decline in ELA test scores for economically disadvantaged students in the first year of exposure, but that marginally significant difference is partially reversed, and becomes insignificant in 2012. Thus, the first cohort appears generally unaffected by this reform, at least in the first two years. Notably, state testing during the study period occurred in the fall and tested standards taught in the previous year, so the lack of effects in 2011 is not unexpected.

The results for the 2012 cohort stand in contrast, however, and indicate that the academic achievement of students in this cohort – particularly economically disadvantaged students – benefitted from these reforms in 2012, the first year they were exposed.

As is typically the case in studies relying on natural experiment techniques, there are some important limitations to acknowledge. We discuss two key limitations here. First, our ability to estimate long-term impacts is limited because our identification strategy uses as the counterfactual the students in districts that have not yet been treated. To ensure a clear counterfactual, our analyses are limited to two outcome years for the 2011 cohort and one outcome year for the 2012 cohort. Thus, while we find some evidence of potential benefits for the second cohort, it is unclear what the longer-term effects are on the 2011 and 2012 cohorts or what the effects are on later cohorts, for whom we do not have a clean counterfactual.

Another limitation concerns the unique aspects of the cohorts on which we are able to estimate effects. In particular, the 2012 cohort, for whom we estimate statistically significant benefits, were somewhat different from the comparison cohorts exposed in 2013 and later. In particular, districts in the 2012 cohort were smaller, were less likely to be urban or suburban, and had relatively restrictive CBAs overall. If some of these district characteristics mediated or moderated reform impacts, the results may not be generalizable to other contexts. In particular, we believe that districts with more restrictive CBAs would be more susceptible to reform, which is consistent with our finding of statistically significant impacts for the 2012 cohort but not the 2011 cohort. This is because the labor market reforms implemented in 2011 served to make overall regulations impacting local teacher markets less restrictive for administrators. This suggests that other areas without strong CBAs and teacher protections may not experience the positive effects on student achievement.

More problematic would be other unobservable factors confounding our estimates. We help address this limitation through the use of student fixed effects to account for time invariant unobservable characteristics of students (and for the vast majority of students, who do not switch districts, this student fixed effect subsumes a district fixed effect as well). In addition, our results are robust to the inclusion of district fixed effects. A remaining concern would exist if there are unobservable shocks to communities or districts that are coinciding with the timing of the reform exposure, and that are affecting treatment and comparison cohorts differentially.

Despite these limitations, the pattern of results adds to the existing evidence on the potential for positive impact of reforms such as rigorous teacher evaluation (Dee & Wyckoff, 2015; Steinberg & Sartain, 2015; Taylor & Tyler, 2012), removal of tenure protections (Carruthers, Figlio, and Sass, 2018), and tying layoff decisions to merit rather than simply seniority (Kraft, 2015; Strunk, Goldhaber, Knight, & Brown, 2018). Similarly, it contributes to the body of work suggesting that stronger unions and collective bargaining may restrict student achievement (e.g., Cowen & Strunk, 2015 Hoxby, 1996; Lovenheim & Willen, 2018; Marianno & Strunk, 2018; Moe, 2009; Strunk, 2011; Strunk & McEachin, 2011).

This leads us to consider the possibility articulated by Goldhaber (2015) in his appraisal of the potential for teacher evaluation to improve education: perhaps what is good for individual teachers does not translate into gains in the educational experience of individual students and, conversely, perhaps what detracts from individual teachers' employment conditions does not by itself translate into another barrier to student success. Indeed, the results from Michigan, suggest, if anything, that the set of labor market reforms, sometimes viewed as a "war on teachers," may indeed be translating into achievement gains for students.

The results of our study here are certainly limited in many ways, but coupled with ongoing work by this team of authors and by others, these results do strengthen an argument to consider policy effects separately for teachers and the students who learn in their classrooms.

References

Baron, E. 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.

Brunner, E., Cowen, J., Strunk, K. & Drake, S. (Forthcoming). Teacher Labor Market Responses to Statewide Reform: Evidence from Michigan. *Educational Evaluation and Policy Analysis.*

Carruthers, C., Figlio, D., & Sass, T. (May 17, 2018). Did tenure reform in Florida affect student test score? *Evidence Speaks Reports, 2*(52). Retrieved 9/25/18 from https://www.brookings.edu/research/did-tenure-reform-in-florida-affect-student-test-scores/

Cowen, J. M., & 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.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT *Journal of Policy Analysis and Management, 34*(2), 267-297.

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher, 44*(2), 87-95.

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.* DOI: 10.1177/2158244013489694

Hambelton, R. K., Swaminathan, J., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: SAGE.

Hoxby, C. M. (1996). How teacher unions effect education production. *Quarterly Journal of Economics, 111,* 671-718.

Janus v. American Federation of State, County, and Municipal Employees, Council 31, et al., 585 U.S. ___ (2018).

Kraft, M. (2015). Teacher layoffs, teacher quality, and student achievement: Evidence from a discretionary layoff policy. *Education Finance and Policy, 10*, 467–507.

Lovenheim, M. F., & Willen, A. (2018). T*he long-run effects of teacher collective bargaining* (No. w24782). National Bureau of Economic Research.

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*(2), 334–359

Marianno, B. & 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*.*

Michigan Association of School Boards (n.d.) *The Bargaining Toolkit: A Resource Manual for School Districts.* Retrieved 9/25/18 from https://www.masb.org/Portals/0/Member_Center/Labor_Relations/Bargaining_Toolkit.pdf

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

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: University of Chicago Press.

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

Roth, J. (March 29, 2019). *Union reform and teacher turnover: Evidence from Wisconsin's Act 10.* Working Paper.

State of Michigan (2011) 96th *Regular Session of 2011 Public Acts 4, 100-103*. Retrieved 10/30/18 from https://www.legislature.mi.gov/documents/2011-2012/publicact/htm/2011-PA-0004.htm http://www.legislature.mi.gov/(S(shpnxyzelc5gk4bdylhw3wuz))/documents/2011-2012/publicact/pdf/2011-PA-0100.pdf ; http://www.legislature.mi.gov/(S(z1qtzld2isehc0wpzfqeppzt))/documents/2011-2012/publicact/pdf/2011-PA-0101.pdf ; https://www.legislature.mi.gov/documents/2011-2012/publicact/htm/2011-PA-0102.htm ; http://www.legislature.mi.gov/(S(f4zpymrzyn3defdbjquqjzwy))/documents/2011-2012/publicact/pdf/2011-pa-0103.pdf

State of Michigan (2012). 96th *Regular Session of 2012 Public Acts 349 and 436.* Retrieved 9/25/18 from https://www.legislature.mi.gov/documents/2011-2012/publicact/htm/2012-PA-0349.htm http://www.legislature.mi.gov/documents/2011-2012/publicact/pdf/2012-PA-0436.pdf

State of Michigan (2015). 98th *Regular Session of 2015 Public Act 173.* Retrieved 9/25/18 from https://www.legislature.mi.gov/documents/2015-2016/publicact/pdf/2015-PA-0173.pdf

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.

Strunk, K. O. (2011). Are teachers' unions really to blame? Collective bargaining agreements and their relationships with district resource allocation and student performance in California. *Education Finance and Policy, 6*(3), 354-398.

Strunk, K. O., Barrett, N., & Lincove, J. A. (2017). When tenure ends: The short-run effects of the elimination of Louisiana's teacher employment protections on teacher exit and retirement. *Education Research Alliance Technical Report*.

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., Goldhaber, D., Knight, D. S., & Brown, N. (2018). Are there hidden costs associated with conducting layoffs? The impact of reduction-in-force and layoff notices on teacher effectiveness. *Journal of Policy Analysis and Management*, *37*(4), 755-782.

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

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.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *The American Economic Review, 102*(7), 3628-3651.

Wright, B., & Masters, G. (1982). *Rating scale analysis: Rasch measurement.* Chicago, IL: MESA Press.

Table 1. Descriptive statistics of analytical samples

	2011 Exposure Cohort - Math Samples	2011 Exposure Cohort - ELA Samples	2012 Exposure Cohort - Math Samples	2012 Exposure Cohort - ELA Samples
Panel A: Excluding Detroit				
Pct. Female	48.78%	49.01%	48.78%	49.04%
Pct. NWNA	18.58%	18.45%	18.23%	18.05%
Pct. Special Needs	12.97%	11.67%	13.56%	12.03%
Pct. LEP	3.44%	3.31%	3.37%	3.24%
Pct. Econ. Dis.	39.73%	39.40%	38.92%	38.51%
Average Grade Level	5.55	5.55	5.83	5.84
Lagged Math Z-Score	0.08	0.08	0.06	0.07
Lagged ELA Z-Score	0.07	0.07	0.06	0.06
Num. Obs.	4,443,167	4,370,489	3,044,411	2,985,729
Panel B: Including Detroit				
	2011 Exposure Cohort - Math Samples	**2011 Exposure Cohort - ELA Samples**	**2012 Exposure Cohort - Math Samples**	**2012 Exposure Cohort - ELA Samples**
Pct. Female	48.81%	49.05%	48.83%	49.10%
Pct. NWNA	22.78%	22.61%	24.57%	24.38%
Pct. Special Needs	13.06%	11.71%	13.64%	12.08%
Pct. LEP	3.70%	3.57%	3.79%	3.66%
Pct. Econ. Dis.	41.96%	41.63%	42.37%	41.98%
Average Grade Level	5.54	5.55	5.81	5.81
Lagged Math Z-Score	0.04	0.05	0.01	0.01
Lagged ELA Z-Score	0.04	0.04	0.01	0.01
Num. Obs.	4,696,576	4,617,128	3,312,615	3,247,410

Note. Demographic characteristics based on percentages or means across all student-year observations. NWNA = Non-white Non-Asian. LEP = Limited English Proficient. Econ. Dis. = economically disadvantaged.

Table 2. 2010 Characteristics and pre-reform CBA restrictiveness of districts, by exposure cohort

	All Districts	Exposed 2011	Exposed 2012	Exposed 2013 and Later
Number of Districts	547	335	148	64
Student Characteristics				
Average Enrollment	1,104	1,144	862	1,453
Average Grade	5.53	5.51	5.57	5.58
% Female	48.4%	48.2%	48.6%	48.8%
% White	82.7%	82.0%	84.1%	82.7%
% Black	7.3%	7.6%	5.5%	9.8%
% Hispanic	5.1%	5.2%	5.5%	3.3%
% Other Race	5.0%	5.2%	4.9%	4.3%
% Econ. Dis.	50.7%	50.6%	51.1%	49.9%
% Special Needs	13.5%	13.4%	13.7%	13.7%
% LEP	1.7%	1.7%	1.7%	1.4%
Avg. Math Z-score	-0.04	-0.04	-0.04	-0.01
Avg. ELA Z-score	0.01	0.00	0.02	0.04
Teacher Characteristics				
Average Num. Teachers	121.5	125.0	93.9	166.8
% Female	73.4%	73.7%	72.9%	72.8%
% White	97.1%	96.8%	97.6%	97.4%
% Black	1.6%	1.9%	1.0%	1.5%
% Hispanic	0.6%	0.6%	0.6%	0.5%
% Other Race	0.7%	0.8%	0.8%	0.6%
% With a Masters +	50.6%	50.0%	49.9%	55.0%
% 1-3 years exp.	8.4%	8.4%	8.7%	7.8%
% 4-6 years exp.	11.5%	11.5%	11.4%	11.8%
% 7-9 years exp.	12.9%	13.3%	12.1%	12.8%
% 10-12 years exp.	17.3%	17.3%	16.7%	18.6%
% 13-15 years exp.	13.6%	13.7%	13.5%	13.1%
% 16-18 years exp.	10.0%	10.0%	10.3%	9.5%
% 19+ years exp.	26.3%	25.8%	27.3%	26.4%
Community Characteristics				
City	6.1%	7.2%	3.4%	6.3%
Suburb	27.6%	26.5%	26.4%	35.9%
Town	17.1%	16.3%	19.6%	15.6%
Rural	48.9%	49.4%	50.7%	42.2%
CBA Restrictiveness				
Overall	-0.03	-0.04	0.04	-0.14
Economic Incentives & Compensation	-0.12	-0.20	0.08	-0.18
Evaluation	-0.01	-0.04	0.01	0.05
Discipline	-0.05	-0.05	-0.05	-0.07
School Days, Hours, & Years	0.05	0.01	0.12	0.10
Transfers and Vacancies	-0.03	-0.02	-0.03	-0.07

Note. 2014 and 2015 exposure cohorts (seven and one districts, respectively) included with the 2013 exposure cohort due to small N. Econ. Dis. = economically disadvantaged status, which in the 2010 year was based solely on FRL-status. LEP = limited English proficient. CBA = collectively bargained agreement.

Table 3. Estimated policy impact on student achievement (2011 exposure cohort, excluding Detroit)

	All Students		Economically Disadvantaged		Non-White Non-Asian	
	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score
Treat X 2007 or Earlier	-0.011	-0.005	0.003	0.002	0.019	0.029**
	(0.017)	(0.009)	(0.015)	(0.011)	(0.017)	(0.013)
Treat X 2008	-0.010	0.005	-0.007	0.005	-0.003	0.000
	(0.014)	(0.008)	(0.013)	(0.009)	(0.018)	(0.012)
Treat X 2009	-0.007	0.004	0.000	0.003	0.011	0.016
	(0.010)	(0.007)	(0.008)	(0.008)	(0.012)	(0.010)
Treat X 2011	-0.001	-0.012	-0.013	-0.020*	-0.019	-0.019
	(0.012)	(0.008)	(0.012)	(0.011)	(0.015)	(0.011)
Treat X 2012	0.034	-0.005	0.019	-0.011	0.020	0.008
	(0.022)	(0.020)	(0.018)	(0.024)	(0.027)	(0.022)
Special Needs	0.042***	-0.003	0.086***	0.011**	0.120***	0.030***
	(0.006)	(0.004)	(0.007)	(0.006)	(0.013)	(0.009)
LEP	-0.052***	-0.067***	-0.031***	-0.058***	-0.011	-0.045***
	(0.013)	(0.009)	(0.009)	(0.010)	(0.008)	(0.010)
Econ. Dis.	-0.009***	-0.006***	-0.002	-0.002	0.007	0.004
	(0.003)	(0.002)	(0.004)	(0.003)	(0.005)	(0.004)
Constant	1.565***	1.646***	1.080***	1.126***	1.061***	1.111***
	(0.080)	(0.085)	(0.067)	(0.071)	(0.108)	(0.117)
Observations	4,443,167	4,370,489	1,771,543	1,727,211	798,222	778,955
R-squared	0.817	0.769	0.762	0.745	0.770	0.758
Adjusted R-squared	0.751	0.686	0.681	0.658	0.686	0.668

Note. Robust standard errors, clustered at the district level, in parentheses. Sample includes students in grades 3-8. 2011 exposure cohort includes districts with collective bargaining agreements (CBAs) that expired in 2009 and 2010 in addition to the districts with CBAs that expired in 2011. Samples for Non-white Non-Asians and Econ. Dis. are based on status in the most recent year, pre-reform, the student was observed. LEP = limited English proficiency. Econ. Dis. = economically disadvantaged. All models include student fixed effects, grade level fixed effects, and school year fixed effects. Excludes Detroit. *** p<0.01, ** p<0.05, * p<0.1

Table 4. Estimated policy impact on student achievement (2012 exposure cohort, excluding Detroit)

	All Students		Economically Disadvantaged		Non-White Non-Asian	
	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score
Treat X 2008 or Earlier	-0.015	-0.005	0.025	0.027	-0.020	0.017
	(0.030)	(0.018)	(0.027)	(0.023)	(0.025)	(0.019)
Treat X 2009	-0.024	-0.006	0.007	0.004	-0.020	-0.023*
	(0.023)	(0.016)	(0.022)	(0.023)	(0.024)	(0.017)
Treat X 2010	-0.019	-0.020	0.005	0.003	-0.017	-0.008
	(0.015)	(0.013)	(0.016)	(0.017)	(0.017)	(0.016)
Treat X 2012	0.043**	0.026*	0.065***	0.058***	0.071***	0.076***
	(0.017)	(0.015)	(0.014)	(0.016)	(0.020)	(0.018)
Special Needs	0.051***	-0.001	0.100***	0.014**	0.138***	0.038***
	(0.007)	(0.004)	(0.009)	(0.006)	(0.015)	(0.010)
LEP	-0.049***	-0.069***	-0.030***	-0.061***	-0.010	-0.048***
	(0.011)	(0.008)	(0.010)	(0.009)	(0.011)	(0.011)
Econ. Dis.	-0.012***	-0.008***	-0.006	-0.005	0.004	0.005
	(0.003)	(0.002)	(0.004)	(0.004)	(0.005)	(0.005)
Constant	2.450***	2.563***	1.806***	1.851***	1.842***	1.853***
	(0.138)	(0.155)	(0.111)	(0.125)	(0.191)	(0.213)
Observations	3,044,411	2,985,729	1,212,688	1,176,287	544,269	528,198
R-squared	0.817	0.774	0.761	0.749	0.769	0.762
Adjusted R-squared	0.749	0.689	0.675	0.658	0.677	0.668

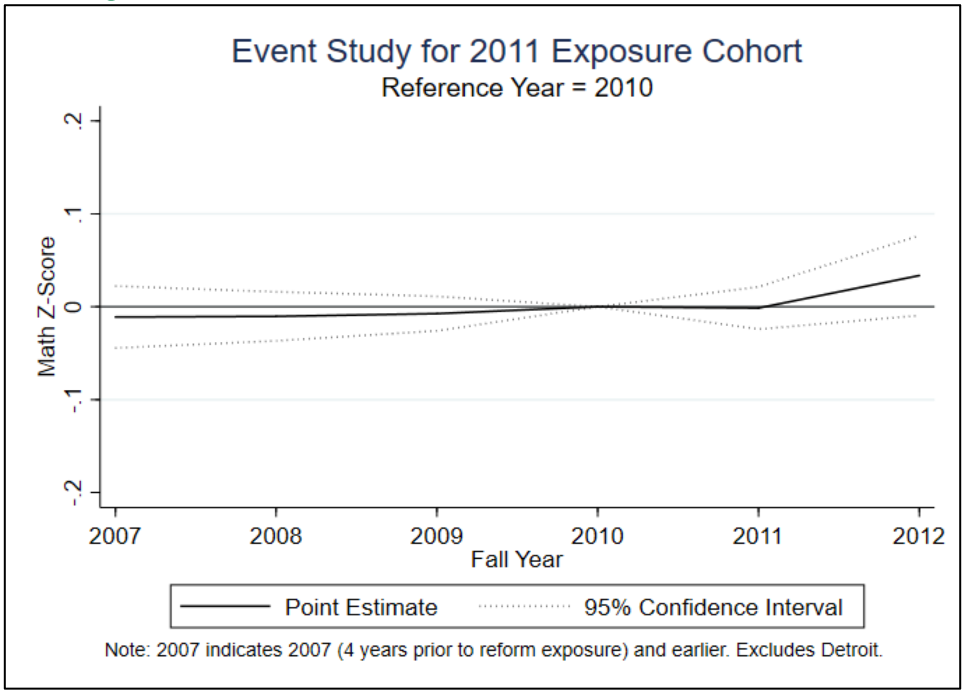
Note. Robust standard errors, clustered at the district level, in parentheses. Sample includes students in grades 3-8. Samples for Non-white Non-Asians and Econ. Dis. are based on status in the most recent year, pre-reform, the student was observed. LEP = limited English proficiency. Econ. Dis. = economically disadvantaged. All models include student fixed effects and grade level fixed effects. Excludes Detroit. *** p<0.01, ** p<0.05, * p<0.1

Figure 1. Timeline of Events.

	2010	2011			2012	2013	2014	2015
Political Timeline	**Nov. 2010:** Gov. Snyder Elected	**Jan. 2011:** Governor Rick Snyder Sworn In	**May 2011:** House Bills 4625- 4628 Proposed (Became PA 100-103)	**July 2011:** PA 100-103 Passed	**Dec. 2012:** PA 349 Passed (Right to Work)	**March 2013:** PA 349 went into effect		**Nov. 2015:** Public Act 173 Passed
Timeline of Reform Exposure				**Fall 2011:** 335 Districts Exposed*	**Fall 2012:** 148 Districts Exposed	**Fall 2013:** 56 Districts Exposed	**Fall 2014:** 7 Districts Exposed	**Fall 2015:** 1 District Exposed

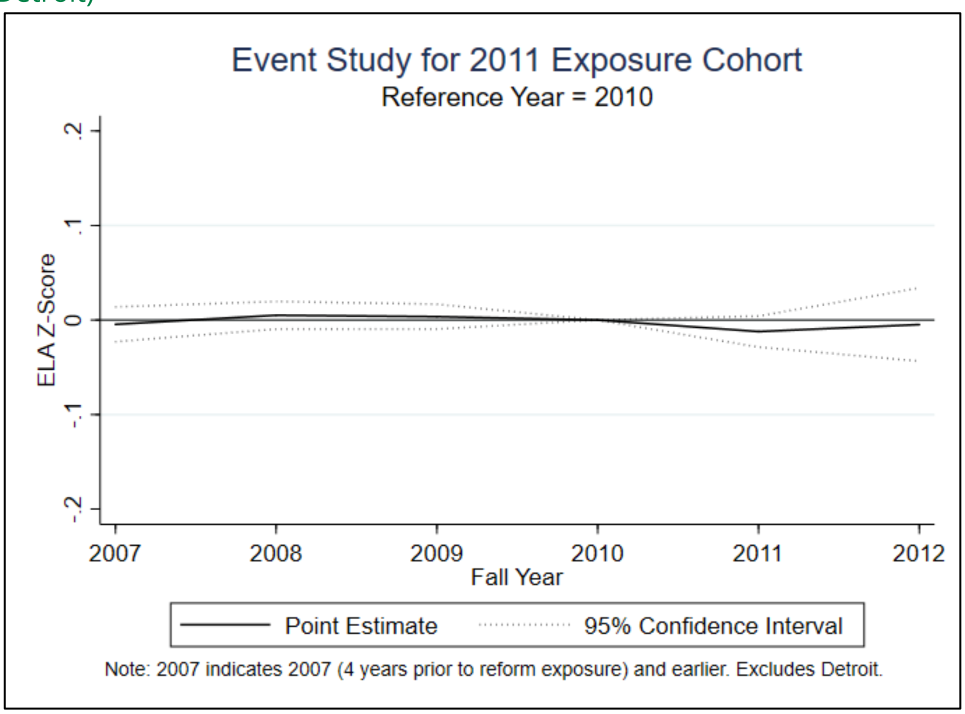
*The 335 districts exposed in fall 2011 include 68 who had CBAs that expired in 2009 and 186 whose CBAs expired in 2010.

Figure 2. Estimated policy impact on math student achievement (2011 exposure cohort, excluding Detroit)



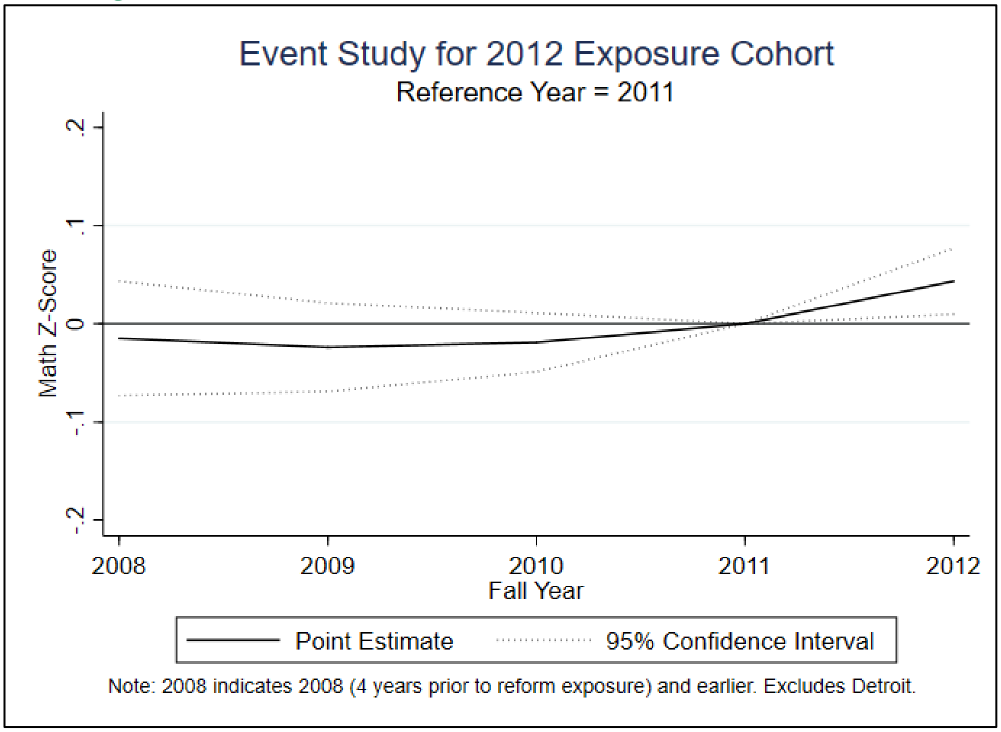
Note. Corresponds to overall math results shown in Table 3. 2011 exposure cohort in this model includes the districts with collective bargaining agreements that expired in 2009 and 2010.

Figure 3. Estimated policy impact on ELA student achievement (2011 exposure cohort, excluding Detroit)



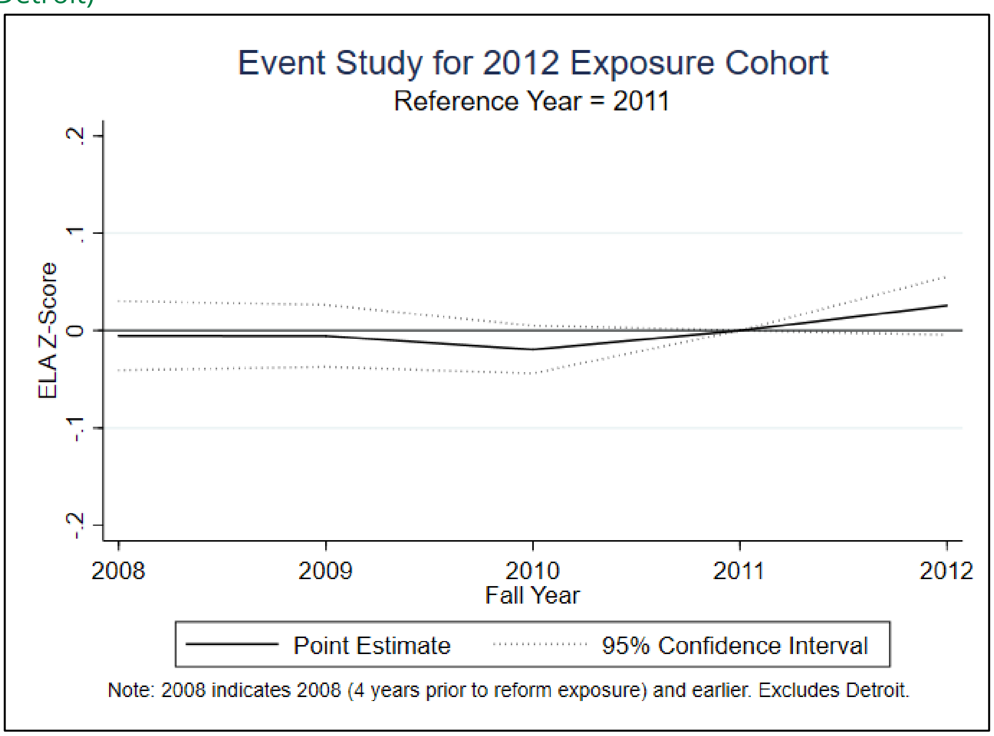
Note. Corresponds to overall ELA results shown in Table 3. 2011 exposure cohort in this model includes the districts with collective bargaining agreements that expired in 2009 and 2010

Figure 4. Estimated policy impact on math student achievement (2012 exposure cohort, excluding Detroit)



Note. Corresponds to overall math results shown in Table 4.

Figure 5. Estimated policy impact on ELA student achievement (2012 exposure cohort, excluding Detroit)



Note. Corresponds to overall ELA results shown in Table 4.

Appendix Table A. Estimated policy impact on student achievement (2011 exposure cohort, excluding Detroit and excluding 2009 and 2010 exposure cohorts)

	All Students		Economically Disadvantaged		Non-White Non-Asian	
	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score
Treat X 2007 or Earlier	0.032	0.008	0.032	0.014	0.044**	0.040**
	(0.024)	(0.014)	(0.023)	(0.017)	(0.019)	(0.017)
Treat X 2008	0.018	0.014	0.011	0.010	0.020	0.007
	(0.018)	(0.012)	(0.020)	(0.016)	(0.022)	(0.016)
Treat X 2009	0.008	0.016	0.012	0.017	0.022	0.024**
	(0.013)	(0.011)	(0.013)	(0.014)	(0.017)	(0.012)
Treat X 2011	-0.010	-0.021	-0.017	-0.027	-0.020	-0.014
	(0.018)	(0.015)	(0.020)	(0.021)	(0.024)	(0.016)
Treat X 2012	0.021	-0.024	0.015	-0.031	0.012	0.008
	(0.030)	(0.032)	(0.027)	(0.043)	(0.036)	(0.030)
Special Needs	0.041***	-0.003	0.087***	0.009	0.121***	0.027**
	(0.007)	(0.004)	(0.009)	(0.007)	(0.016)	(0.011)
LEP	-0.048***	-0.062***	-0.028***	-0.053***	-0.013	-0.041***
	(0.016)	(0.011)	(0.010)	(0.012)	(0.009)	(0.012)
Econ. Dis.	-0.009**	-0.008***	0.000	-0.004	0.007	-0.002
	(0.004)	(0.003)	(0.005)	(0.005)	(0.006)	(0.006)
Constant	1.967***	2.040***	1.413***	1.452***	1.340***	1.391***
	(0.130)	(0.146)	(0.109)	(0.125)	(0.164)	(0.189)
Observations	3,066,020	3,007,471	1,178,224	1,142,909	572,803	556,282
R-squared	0.826	0.779	0.766	0.752	0.771	0.762
Adjusted R-squared	0.757	0.691	0.680	0.660	0.681	0.668

Note. Robust standard errors, clustered at the district level, in parentheses. Sample includes students in grades 3-8. 2011 exposure cohort in this model excludes the districts with collective bargaining agreements that expired in 2009 and 2010. Samples for Non-white Non-Asians and Econ. Dis. are based on status in the most recent year, pre-reform, the student was observed. LEP = limited English proficiency. Econ. Dis. = economically disadvantaged. All models include student fixed effects and grade level fixed effects. Excludes Detroit. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table B. Estimated policy impact on student achievement (Full sample, excluding Detroit, moving reference year one year back)

	2011 Cohort (Including 2009 and 2010 Cohorts)		2011 Cohort (Excluding 2009 and 2010 Cohorts)			2012 Cohort	
	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score		Math Z-score	ELA Z-score
Treat X 2007 or Earlier	-0.004	-0.008	0.024	-0.008	Treat X 2008 or Earlier	0.004	0.014
	(0.014)	(0.006)	(0.021)	(0.008)		(0.021)	(0.010)
Treat X 2008	-0.003	0.001	0.010	-0.002	Treat X 2009	-0.005	0.014*
	(0.008)	(0.005)	(0.011)	(0.007)		(0.012)	(0.008)
Treat X 2010	0.007	-0.004	-0.008	-0.016			
	(0.010)	(0.007)	(0.013)	(0.011)			
Treat X 2011	0.006	-0.016	-0.018	-0.037	Treat X 2011	0.019	0.020
	(0.018)	(0.013)	(0.026)	(0.024)		(0.015)	(0.013)
Treat X 2012	0.041	-0.008	0.013	-0.039	Treat X 2012	0.062**	0.045**
	(0.028)	(0.024)	(0.037)	(0.040)		(0.024)	(0.022)
Special Needs	0.042***	-0.003	0.041***	-0.003	Special Needs	0.051***	-0.001
	(0.006)	(0.004)	(0.007)	(0.004)		(0.007)	(0.004)
LEP	-0.052***	-0.067***	-0.048***	-0.062***	LEP	-0.049***	-0.069***
	(0.013)	(0.009)	(0.016)	(0.011)		(0.011)	(0.008)
Econ. Dis.	-0.009***	-0.006***	-0.009**	-0.008***	Econ. Dis.	-0.012***	-0.008***
	(0.003)	(0.002)	(0.004)	(0.003)		(0.003)	(0.002)
Constant	1.211***	1.264***	1.566***	1.612***	Constant	2.073***	2.166***
	(0.062)	(0.065)	(0.103)	(0.115)		(0.115)	(0.130)
Observations	4,443,167	4,370,489	3,066,020	3,007,471	Observations	3,044,411	2,985,729
R-squared	0.817	0.769	0.826	0.779	R-squared	0.817	0.774
Adjusted R-squared	0.751	0.686	0.757	0.691	Adjusted R-squared	0.749	0.689

Note. Robust standard errors, clustered at the district level, in parentheses. Sample includes students in grades 3-8. Samples for Non-white Non-Asians and Econ. Dis. are based on status in the most recent year, pre-reform, the student was observed. LEP = limited English proficiency. Econ. Dis. = economically disadvantaged. All models include student fixed effects and grade level fixed effects. Excludes Detroit. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table C. Estimated policy impact on student achievement (2011 exposure cohort, including Detroit)

	All Students		Economically Disadvantaged		Non-White Non-Asian	
	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score
Treat X 2007 or Earlier	-0.012	-0.013	0.007	-0.006	0.005	-0.002
	(0.017)	(0.014)	(0.014)	(0.015)	(0.020)	(0.031)
Treat X 2008	-0.010	0.000	-0.003	0.000	-0.004	-0.018
	(0.013)	(0.010)	(0.012)	(0.011)	(0.016)	(0.019)
Treat X 2009	-0.009	-0.003	0.001	-0.005	0.005	-0.009
	(0.009)	(0.010)	(0.008)	(0.012)	(0.012)	(0.022)
Treat X 2011	0.004	-0.009	-0.005	-0.013	0.004	-0.001
	(0.012)	(0.009)	(0.013)	(0.013)	(0.021)	(0.015)
Treat X 2012	0.036	-0.005	0.023	-0.011	0.029	0.008
	(0.022)	(0.020)	(0.018)	(0.024)	(0.027)	(0.022)
Special Needs	0.047***	-0.004	0.092***	0.010*	0.130***	0.030***
	(0.008)	(0.004)	(0.010)	(0.005)	(0.015)	(0.008)
LEP	-0.056***	-0.069***	-0.038***	-0.063***	-0.024**	-0.052***
	(0.013)	(0.009)	(0.011)	(0.010)	(0.012)	(0.009)
Econ. Dis.	-0.005	-0.001	0.000	0.004	0.014***	0.017**
	(0.004)	(0.006)	(0.004)	(0.006)	(0.005)	(0.008)
Constant	2.358***	2.381***	1.944***	1.954***	2.540***	2.564***
	(0.577)	(0.526)	(0.598)	(0.568)	(0.660)	(0.638)
Observations	4,696,576	4,617,128	1,974,433	1,924,583	1,043,304	1,017,736
R-squared	0.812	0.767	0.748	0.736	0.734	0.734
Adjusted R-squared	0.744	0.683	0.660	0.644	0.633	0.634

Note. Robust standard errors, clustered at the district level, in parentheses. Sample includes students in grades 3-8. 2011 exposure cohort in this model includes the districts with collective bargaining agreements that expired in 2009 and 2010, in addition to the districts with CBAs that expired in 2011. Samples for Non-white Non-Asians and Econ. Dis. are based on status in the most recent (CBAs) year, pre-reform, the student was observed. LEP = limited English proficiency. Econ. Dis. = economically disadvantaged. All models include student fixed effects and grade level fixed effects. Includes Detroit. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table D. Estimated policy impact on student achievement (2012 exposure cohort, including Detroit)

	All Students		Economically Disadvantaged		Non-White Non-Asian	
	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score
Treat X 2008 or Earlier	-0.001	0.013	0.038	0.050*	0.035	0.066**
	(0.030)	(0.025)	(0.026)	(0.028)	(0.037)	(0.026)
Treat X 2009	-0.007	0.002	0.030	0.014	0.040	0.002
	(0.027)	(0.017)	(0.027)	(0.022)	(0.039)	(0.014)
Treat X 2010	-0.009	-0.012	0.018	0.012	0.017	0.008
	(0.017)	(0.014)	(0.017)	(0.016)	(0.020)	(0.012)
Treat X 2012	0.044***	0.024	0.068***	0.050***	0.075***	0.059***
	(0.017)	(0.015)	(0.013)	(0.017)	(0.017)	(0.020)
Special Needs	0.060***	0.000	0.109***	0.017**	0.150***	0.045***
	(0.012)	(0.005)	(0.014)	(0.007)	(0.017)	(0.011)
LEP	-0.053***	-0.070***	-0.038***	-0.064***	-0.027*	-0.053***
	(0.011)	(0.008)	(0.012)	(0.010)	(0.014)	(0.009)
Econ. Dis.	-0.006	0.000	-0.003	0.002	0.013***	0.020***
	(0.006)	(0.007)	(0.004)	(0.007)	(0.005)	(0.008)
Constant	3.601***	3.641***	3.075***	3.093***	3.701***	3.734***
	-0.747	-0.691	-0.782	-0.759	-0.706	-0.704
Observations	3,312,615	3,247,410	1,420,759	1,379,037	795,520	773,395
R-squared	0.809	0.770	0.740	0.734	0.720	0.728
Adjusted-Rsquared	0.738	0.684	0.646	0.638	0.609	0.620

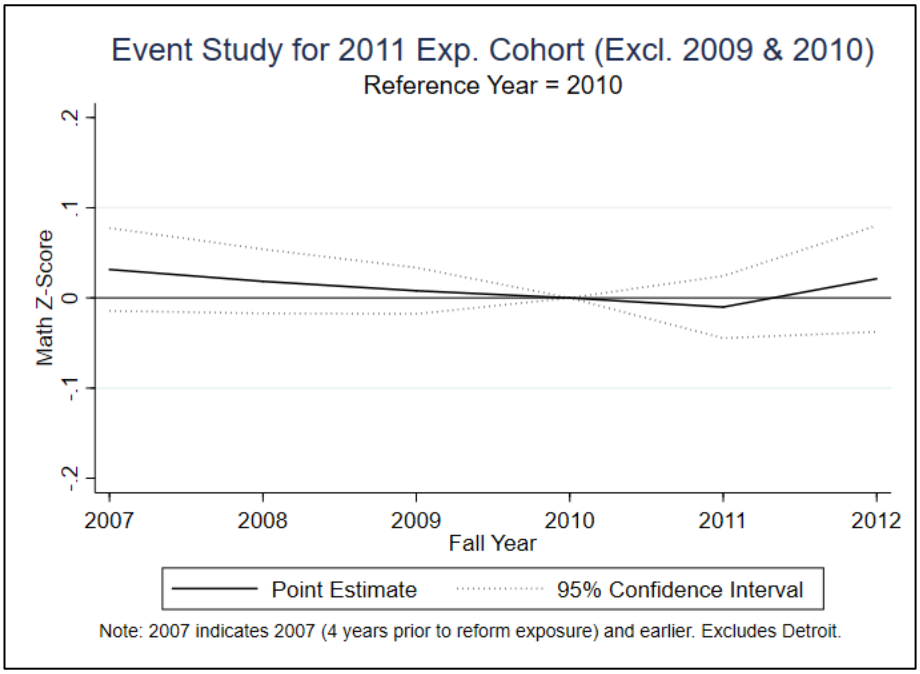
Note. Robust standard errors, clustered at the district level, in parentheses. Sample includes students in grades 3-8. Samples for Non-white Non-Asians and Econ. Dis. are based on status in the most recent year, pre-reform, the student was observed. LEP = limited English proficiency. Econ. Dis. = economically disadvantaged. All models include student fixed effects and grade level fixed effects. Includes Detroit. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table E. Estimated policy impact on student achievement (2011 exposure cohort, including Detroit and excluding 2009 and 2010 exposure cohorts)

	All Students		Economically Disadvantaged		Non-White Non-Asian	
	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score	Math Z-score	ELA Z-score
Treat X 2007 or Earlier	0.026	-0.010	0.036*	-0.003	0.028	-0.005
	(0.024)	(0.025)	(0.019)	(0.025)	(0.023)	(0.038)
Treat X 2008	0.016	0.002	0.017	-0.001	0.019	-0.020
	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)	(0.024)
Treat X 2009	0.005	0.003	0.014	0.002	0.018	-0.010
	(0.013)	(0.017)	(0.011)	(0.019)	(0.015)	(0.027)
Treat X 2011	0.002	-0.012	0.001	-0.009	0.014	0.009
	(0.020)	(0.016)	(0.022)	(0.021)	(0.024)	(0.014)
Treat X 2012	0.029	-0.021	0.023	-0.030	0.027	0.007
	(0.030)	(0.032)	(0.027)	(0.043)	(0.035)	(0.031)
Special Needs	0.049***	-0.004	0.096***	0.008	0.133***	0.029***
	(0.011)	(0.004)	(0.014)	(0.006)	(0.018)	(0.009)
LEP	-0.054***	-0.065***	-0.037***	-0.059***	-0.026**	-0.046***
	(0.016)	(0.011)	(0.013)	(0.014)	(0.012)	(0.011)
Econ. Dis.	-0.003	0.000	0.003	0.004	0.015***	0.015*
	(0.006)	(0.008)	(0.005)	(0.007)	(0.005)	(0.009)
Constant	2.955***	2.967***	2.514***	2.521***	2.994***	3.033***
	(0.652)	(0.603)	(0.682)	(0.659)	(0.650)	(0.635)
Observations	3,312,084	3,246,858	1,374,750	1,334,006	811,233	788,501
R-squared	0.818	0.775	0.746	0.739	0.727	0.733
Adjusted R-squared	0.746	0.686	0.651	0.640	0.617	0.625

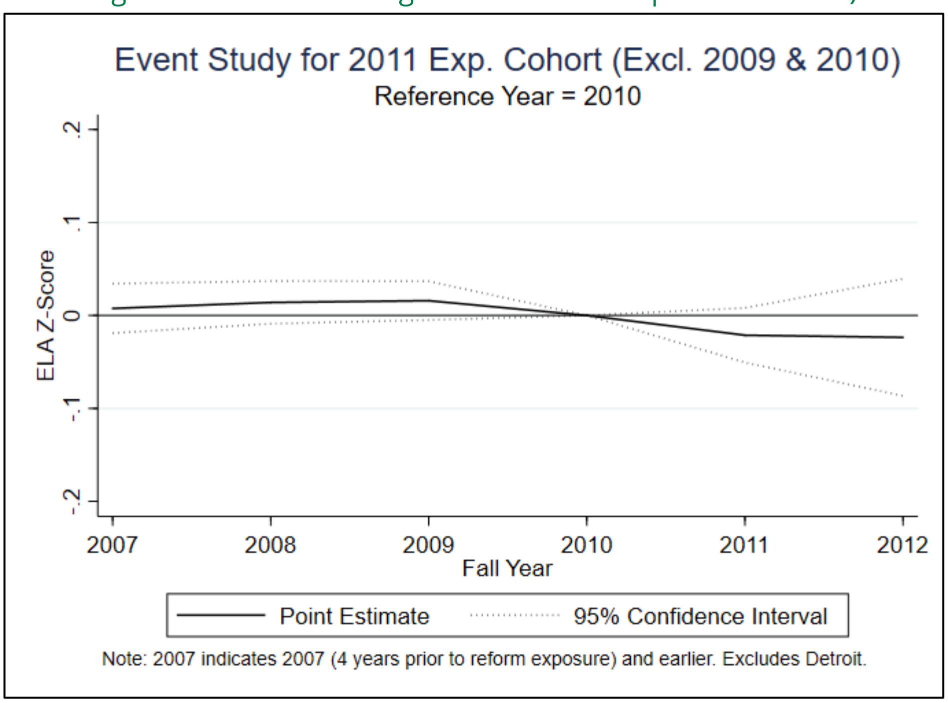
Note. Robust standard errors, clustered at the district level, in parentheses. Sample includes students in grades 3-8. 2011 exposure cohort in this model excludes the districts with collective bargaining agreements that expired in 2009 and 2010. Samples for Non-white Non-Asians and Econ. Dis. are based on status in the most recent year, pre-reform, the student was observed. LEP = limited English proficiency. Econ. Dis. = economically disadvantaged. All models include student fixed effects and grade level fixed effects. Includes Detroit. *** p<0.01, ** p<0.05, * p<0.1

Appendix Figure A. Estimated policy impact on math student achievement (2011 exposure cohort, excluding Detroit and excluding 2009 and 2010 exposure cohorts)



Note. Corresponds to overall math results shown in Appendix Table A. Excludes the districts with collective bargaining agreements that expired in 2009 and 2010.

Appendix Figure B. Estimated policy impact on ELA student achievement (2011 exposure cohort, excluding Detroit and excluding 2009 and 2010 exposure cohorts)



Note. Corresponds to overall ELA results shown in Appendix Table A. Excludes the districts with collective bargaining agreements that expired in 2009 and 2010.